KNN_CLASSIFIER

```
#Aim: To Perform Operation on SVM Classifier
# Name : Samruddhi Umap
# Roll no : 67
# Sec: C
# Subject : ET1
# Date :27/09/2024
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')
import os
os.getcwd()
'C:\\Users\\HP'
os.chdir('C:\\Users\\HP\\DESKTOP')
df=pd.read_csv("framingham.csv")
df.head()
   male age education currentSmoker cigsPerDay
prevalentStroke \
                    4.0
0
      1
          39
                                                0.0
                                                        0.0
0
1
      0
          46
                    2.0
                                               0.0
                                                        0.0
0
2
      1
          48
                    1.0
                                              20.0
                                                        0.0
0
3
      0
                    3.0
                                              30.0
                                                        0.0
          61
0
4
      0
          46
                    3.0
                                     1
                                              23.0
                                                        0.0
   prevalentHyp diabetes totChol sysBP diaBP
                                                    BMI heartRate
glucose \
              0
                             195.0
                                    106.0
                                            70.0 26.97
                                                               80.0
77.0
              0
                        0
                             250.0 121.0
                                            81.0 28.73
                                                               95.0
76.0
```

2 70.0	0	0	245.0	127.5	80.0	25.34	1	75.0
3	1	0	225.0	150.0	95.0	28.58	3	65.0
103.0		_						
4 85.0	0	0	285.0	130.0	84.0	23.10	9)	85.0
03.0								
	arCHD							
0 1	0 0							
1 2 3 4	Õ							
3	1							
4	0							
df.tail()							
	le age	education			cigsPe	•	BPMeds	\
4233 4234	1 50 1 51	1.0 3.0		1 1		1.0 43.0	0.0	
4234 4235	1 51 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	
pr	evalentS	troke pre	valentHyp	diabe	tes to	tChol	sysBP	diaBP
BMI \		0	1		0	212 0	170.0	02.0
4233 25.97		0	1		0	313.0	179.0	92.0
4234		0	0		0	207.0	126.5	80.0
19.71		0	0		0	240 0	121 0	72.0
4235 22.00		0	0		0	248.0	131.0	72.0
4236		0	0		0	210.0	126.5	87.0
19.16		0	0		0	260.0	122 5	02.0
4237 21.47		0	0		0	269.0	133.5	83.0
	artRate	_	TenYearCHD					
4233 4234	66.0 65.0	86.0 68.0	1 6					
4235	84.0	86.0	e)				
4236	86.0	NaN	0					
4237	80.0	107.0	6					
df.shape								
(4238, 1	6)							
df.size								
67808								

df.in	fo											
					of BPMeds		male	age	e educat	ion		
0 1	1	39	J	4.0		•	0 0		0.0 0.0	0.0 0.0		
2	1			1.0			1		20.0	0.0		
3	0			3.0			1		30.0	0.0		
4	0	46		3.0			1		23.0	0.0		
4222		 50		1 0					1.0			
4233 4234	1 1			1.0 3.0			1 1		1.0 43.0	$0.0 \\ 0.0$		
4235	0			2.0			1		20.0	NaN		
4236	0			1.0			1		15.0	0.0		
4237	0	52		2.0			0		0.0	0.0		
BMI	prev	alents	Stroke	preva	ılentHyp)	diabet	es	totChol	sysBP	diaBP	
0 26.97	•		0		(9		0	195.0	106.0	70.0	
1			0		(9		0	250.0	121.0	81.0	
28.73			0		(9		0	245.0	127.5	80.0	
25.34 3			0			1		0	225.0	150.0	95.0	
28.58 4			0		(9		0	285.0	130.0	84.0	
23.10			U		,	J		U	203.0	130.0	04.0	
						•		• •				
4233			0		-	1		0	313.0	179.0	92.0	
25.97 4234			0		(9		0	207.0	126.5	80.0	
19.71			-									
4235 22.00			0		(9		0	248.0	131.0	72.0	
4236			0		(9		0	210.0	126.5	87.0	
19.16 4237			0		(9		0	269.0	133.5	83.0	
21.47			0		,	,		U	203.0	133.3	05.0	
	hear	tRate	gluco	se Te	nYearCl	HD						
0		80.0	77	. 0		0						
1		95.0	76			0						
2		75.0	70 102			0						
2 3 4		65.0 85.0	103 85			0						
4000					• :							
4233 4234		66.0 65.0	86 68			1						
7237		05.0	00	. 0		U						

4236	84.0 86.0 80.0	86.0 NaN 107.0	0 0 0		
[4238 rows	x 16 col	umns]>			
df.describe	()				
cigsPerDay	male	age	education	currentSmoker	
count 4238	.000000	4238.000000	4133.000000	4238.000000	
	.429212	49.584946	1.978950	0.494101	
	.495022	8.572160	1.019791	0.500024	
	.000000	32.000000	1.000000	0.000000	
	.000000	42.000000	1.000000	0.000000	
	.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					
totChol \	BPMeds	prevalentStro	ke prevalen	tHyp diabe	tes
count 4185 4188.000000		4238.0000	00 4238.000	9000 4238.000	000
	.029630	0.0058	99 0.310	0524 0.025	720
	.169584	0.0765	87 0.462	2763 0.158	316
min 0	.000000	0.0000	00 0.000	0.000	000
	.000000	0.0000	00 0.00	0.000	000
	.000000	0.0000	00 0.000	0.000	000
	.000000	0.0000	00 1.000	0.000	000
	.000000	1.0000	00 1.000	9000 1.000	000
696.000000					
\	sysBP	diaBP	BMI	heartRate	glucose
count 4238	.000000	4238.000000	4219.000000	4237.000000	3850.000000

mean	132.352	2407	82.893464	25.8	02008	75.8789	924 8	1.966753
std	22.038	8097	11.910850	4.0	80111	12.0265	596 2	3.959998
min	83.500	0000	48.000000	15.5	40000	44.0000	000 4	0.000000
25%	117.000	0000	75.000000	23.0	70000	68.0000	000 7	1.000000
50%	128.000	0000	82.000000	25.4	00000	75.0006	000 7	8.000000
75%	144.000	0000	89.875000	28.0	40000	83.0006	000 8	7.000000
max	295.000	0000 1	42.500000	56.8	00000	143.0000	000 39	4.000000
count mean std min 25% 50% 75% max df	TenYear 4238.000 0.151 0.359 0.000 0.000 0.000 1.000	9000 958 9023 9000 9000						
0 1 2 3 4 4233 4234 4235 4236 4237	male age 1 39 0 46 1 48 0 61 0 46 1 50 1 51 0 48 0 44 0 52		tion curre 4.0 2.0 1.0 3.0 3.0 1.0 3.0 2.0 1.0	ntSmok	er cig 0 0 1 1 1 1 1 1 0	sPerDay 0.0 0.0 20.0 30.0 23.0 1.0 43.0 20.0 15.0 0.0	NaN	
	•	:Stroke	prevalentH	lyp di	abetes	totChol	sysBP	diaBP
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
		-						

25.97 4234	23.10							
4233								
## 4234	4233		0	1	0	313.0	179.0	92.0
19.71 4235			0	Θ	Θ	207.0	126.5	80.0
22.00 4236	19.71							
4236			Θ	Θ	Θ	248.0	131.0	/2.0
### 133.5 ### 13	4236		0	0	0	210.0	126.5	87.0
heartRate glucose TenYearCHD 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] df.isna().sum() 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 ddype: int64	4237		Θ	0	Θ	269.0	133.5	83.0
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] df.isna().sum() male 0 0 age 0 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 3888 TenYearCHD 0 dtype: int64	21.47							
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] df.isna().sum() male 0 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	0 1 2 3 4	80.0 95.0 75.0 65.0	77.0 76.0 70.0 103.0	0 0 0 1				
df.isna().sum() 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	4233 4234 4235 4236 4237	66.0 65.0 84.0 86.0	86.0 68.0 86.0 NaN	1 0 0 0				
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	[4238	rows x 16	columns]					
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	df.isr	na(). <mark>sum</mark> ()						
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		()	0					
BMI 19 heartRate 1 glucose 388 TenYearCHD 0 dtype: int64	age educat currer cigsPe BPMeds preval preval diabet totChe sysBP	ntSmoker erDay s lentStroke lentHyp tes	0 105 0 29 53 0 0 0 50					
df.isnull()	BMI heart glucos TenYea	se arCHD	19 1 388					
	df.isr	null()						

0 1 2 3 4	male False False False False False	age False False False False	education False False False False		ntSmoker False False False False	cigsPerDa Fals Fals Fals Fals	e Fal e Fal e Fal e Fal e Fal	se se se
4233 4234 4235 4236 4237	False False False False False	False False False False False	False False False False		False False False False False	Fals Fals Fals Fals Fals	e Fal e Fal e Tr e Fal	se se ue se
DMT		entStro	ke preval	.entHyp	diabetes	totChol	sysBP	diaBP
0	\	Fal	se	False	False	False	False	False
False 1		Fal	se	False	False	False	False	False
False		Fal	se	False	False	False	False	False
False		Fal	se	False	False	False	False	False
False		Fal	se	False	False	False	False	False
False								
4233		Fal	se	False	False	False	False	False
False 4234		Fal	se	False	False	False	False	False
False 4235		Fal	se	False	False	False	False	False
False 4236		Fal	se	False	False	False	False	False
False 4237		Fal	se	False	False	False	False	False
False								
0 1 2 3 4 4233 4234 4235 4236 4237	Fa Fa Fa Fa Fa Fa Fa	lse lse lse lse lse lse lse	ucose Ter False False False False False False True False	YearCHD False False False False False False False False				

```
[4238 rows \times 16 columns]
df.isnull().any()
                    False
male
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
                     True
glucose
TenYearCHD
                    False
dtype: bool
```

MISSING VALUE TREATMENT

```
df['qlucose'].fillna(value = df['qlucose'].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)
df.isna().sum()
male
                   0
                   0
age
                   0
education
currentSmoker
                   0
cigsPerDay
                   0
BPMeds
                   0
prevalentStroke
                   0
prevalentHyp
                   0
diabetes
                   0
                   0
totChol
                   0
sysBP
diaBP
                   0
BMT
                   0
heartRate
                   0
```

```
glucose
                    0
TenYearCHD
                    0
dtype: int64
#Splitting the dependent and independent variables.
x = df.drop("TenYearCHD",axis=1)
y = df['TenYearCHD']
Х
      male
                  education currentSmoker
                                              cigsPerDay
                                                            BPMeds \
             age
0
             39
                        4.0
                                                           0.00000
                                                     0.0
         1
                                           0
                        2.0
1
         0
             46
                                           0
                                                     0.0
                                                           0.00000
2
         1
             48
                        1.0
                                           1
                                                    20.0
                                                           0.00000
3
              61
                        3.0
                                           1
                                                    30.0
                                                           0.00000
         0
4
                                           1
                                                    23.0
         0
             46
                        3.0
                                                          0.00000
             . . .
. . .
                                                           0.00000
4233
         1
             50
                        1.0
                                           1
                                                     1.0
4234
             51
                        3.0
                                           1
                                                    43.0
                                                           0.00000
         1
4235
             48
                                           1
         0
                        2.0
                                                    20.0
                                                          0.02963
4236
              44
                        1.0
                                           1
                                                    15.0
                                                           0.00000
         0
              52
4237
         0
                        2.0
                                           0
                                                     0.0
                                                          0.00000
      prevalentStroke prevalentHyp diabetes totChol
                                                            sysBP
                                                                   diaBP
BMI \
                     0
                                                    195.0
                                                            106.0
                                                                    70.0
26.97
                     0
                                    0
                                                    250.0
                                                            121.0
                                                                    81.0
28.73
                     0
                                    0
                                                    245.0
                                                           127.5
                                                                    80.0
2
25.34
                     0
                                                    225.0
                                                            150.0
                                                                    95.0
3
28.58
                                                    285.0
                                                            130.0
                                                                    84.0
23.10
. . .
4233
                                                    313.0
                                                            179.0
                                                                    92.0
25.97
4234
                                    0
                                                    207.0
                                                            126.5
                                                                    80.0
19.71
4235
                     0
                                    0
                                                    248.0
                                                            131.0
                                                                    72.0
22.00
4236
                     0
                                    0
                                                                    87.0
                                                    210.0
                                                           126.5
19.16
                     0
4237
                                                    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]
#Splitting the dependent and independent variables.
x = df.drop("TenYearCHD",axis=1)
v = df['TenYearCHD']
Χ
                  education currentSmoker
                                               cigsPerDay
                                                             BPMeds \
      male
             age
                         4.0
0
          1
              39
                                           0
                                                      0.0
                                                            0.00000
1
                         2.0
          0
              46
                                            0
                                                      0.0
                                                            0.00000
2
              48
                         1.0
                                            1
                                                     20.0
          1
                                                            0.00000
3
              61
                         3.0
                                            1
                                                     30.0
                                                            0.00000
          0
4
              46
                         3.0
                                            1
                                                     23.0
                                                            0.00000
          0
              50
                                                      1.0
                                                            0.00000
4233
         1
                         1.0
                                            1
4234
              51
                         3.0
                                            1
                                                     43.0
         1
                                                            0.00000
4235
              48
                                            1
                                                     20.0
                                                            0.02963
                         2.0
          0
4236
              44
                                            1
                                                     15.0
                                                            0.00000
          0
                         1.0
4237
          0
              52
                         2.0
                                            0
                                                      0.0
                                                            0.00000
                       prevalentHyp
                                        diabetes
      prevalentStroke
                                                   totChol
                                                             sysBP
                                                                     diaBP
BMI \
                      0
                                     0
                                                0
                                                     195.0
                                                            106.0
                                                                      70.0
0
26.97
                      0
                                     0
                                                     250.0
                                                             121.0
                                                                      81.0
1
28.73
                      0
                                                     245.0
                                                             127.5
                                                                      80.0
25.34
                      0
                                                     225.0
                                                             150.0
                                                                      95.0
28.58
                      0
                                     0
                                                     285.0
                                                             130.0
                                                                      84.0
23.10
. . .
. . .
                      0
                                                     313.0
                                                             179.0
                                                                      92.0
4233
25.97
4234
                      0
                                                     207.0 126.5
                                                                      80.0
19.71
```

```
4235
                     0
                                    0
                                                   248.0
                                                           131.0
                                                                   72.0
22.00
4236
                     0
                                    0
                                                   210.0
                                                           126.5
                                                                   87.0
19.16
4237
                     0
                                                   269.0 133.5
                                                                   83.0
21.47
      heartRate
                     glucose
0
                   77.000000
           80.0
1
           95.0
                  76.000000
2
           75.0
                  70.000000
3
                 103.000000
           65.0
4
           85.0
                  85.000000
                  86,000000
4233
           66.0
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]
x_train,x_test,y_train,y_test =
train_test_split(x,y,test_size=0.2,random_state=42)
y_train
3252
        0
3946
        0
1261
        0
2536
        0
4089
        0
3444
        0
        0
466
3092
        0
3772
        0
860
Name: TenYearCHD, Length: 3390, dtype: int64
```

KNN

```
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
```