KNN_CLASSIFIER

```
#Aim: To Perform Operation on SVM Classifier
# Name : Achal Chandure
# Roll no : 08
# 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 BPMeds
prevalentStroke \
                    4.0
                                               0.0
                                                       0.0
  1 39
0
1
  0 46
                    2.0
                                              0.0
                                                       0.0
0
2
  1 48
                    1.0
                                              20.0
                                                      0.0
0
3
      0
          61
                    3.0
                                              30.0
                                                       0.0
0
4
                    3.0
      0
        46
                                              23.0
                                                       0.0
  prevalentHyp diabetes totChol sysBP diaBP BMI heartRate
glucose \
                          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		75.0
3	1	0	225.0	150.0	95.0	28.58		65.0
103.0	0	0	285.0	130.0	84.0	23.10		85.0
85.0	U	0	203.0	130.0	04.0	23.10		03.0
TenYe 0 1 2 3 4	earCHD 0 0 0 1							
df.tail	()							
4233 4234 4235 4236 4237	ale age 1 50 1 51 0 48 0 44 0 52	education 1.0 3.0 2.0 1.0 2.0	currentS	moker 1 1 1 1 0	cigsP	1.0 43.0 20.0 15.0 0.0	BPMeds 0.0 0.0 NaN 0.0	\
BMI \	revalentS	troke preva	alentHyp	diabet	tes t	otChol	sysBP	diaBP
4233		0	1		0	313.0	179.0	92.0
25.97 4234		0	0		0	207.0	126.5	80.0
19.71		U	U		U	207.0	120.5	00.0
4235 22.00		0	0		0	248.0	131.0	72.0
4236		0	0		0	210.0	126.5	87.0
19.16 4237		0	0		0	269.0	133.5	83.0
21.47		U	U		U	209.0	133.3	03.0
ho 4233 4234 4235 4236 4237	eartRate 66.0 65.0 84.0 86.0 80.0	glucose T 86.0 68.0 86.0 NaN 107.0	enYearCHD 1 0 0 0					
df.shape	9							
(4238, 1	₋ 6)							
df.size								
67808								

df.inf	T O							
currer	ntSmoker c	igsPerDay			age			
0	1 39 0 46	4.0 2.0		0 0		0.0	0.0	
2	1 48 0 61	1.0		1 1		20.0 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 4236	0 48 0 44	2.0		1 1		20.0 15.0	NaN 0.0	
4237	0 52	2.0		0		0.0	0.0	
	prevalen							diaBP
BMI \		0	0		0	105.0	1000	70 0
0 26.97		0	U		U	195.0	106.0	70.0
1 28.73		0	0		0	250.0	121.0	81.0
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 23.10		0	0		0	285.0	130.0	84.0
			• • •			• • •		
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 22.00		0	0		0	248.0	131.0	72.0
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 1	80.0 95.0	77.0 76.0	0					
2	75.0	70.0	0					
3 4	65.0 85.0	103.0 85.0	1 0					
4233 4234	66.0 65.0	86.0 68.0	1 0					

4235 4236 4237	84.0 86.0 80.0	86.0 NaN 107.0	0 0 0			
[4238 rows	s x 16 col	umns]>				
df.describ	pe()					
	male	age	educa	ation c	urrentSmoker	
cigsPerDay count 423 4209.00000	38.000000	4238.000000	4133.00	0000	4238.00000	0
mean 9.003089	0.429212	49.584946	1.9	78950	0.49410	1
std 11.920094	0.495022	8.572160	1.0	19791	0.50002	4.4
min 0.000000	0.000000	32.000000	1.0	00000	0.00000	0
25% 0.000000	0.000000	42.000000	1.00	00000	0.00000	0
50%	0.000000	49.000000	2.0	00000	0.00000	0
75% 20.000000	1.000000	56.000000	3.00	00000	1.00000	0
max 70.000000	1.000000	70.000000	4.00	00000	1.00000	0
	D DM a d a	10 m + C+	l-			0.1.0.0
totChol \	BPMeas	prevalentStro	ke prev	arenthy	p diab	etes
count 418 4188.00000		4238.000	000 42	238.0000	000 4238.00	0000
mean 236.721585	0.029630	0.005	899	0.3105	524 0.02	5720
std 44.590334	0.169584	0.076	587	0.4627	763 0.15	8316
min 107.000000	0.000000	0.000	000	0.0000	0.00	0000
25% 206.000000	0.000000	0.000	000	0.0000	0.00	0000
50% 234.000000	0.000000	0.000	000	0.0000	0.00	0000
75% 263.000000	0.000000	0.000	000	1.0000	0.00	0000
max 696.000000	1.000000	1.000	000	1.0000	1.00	0000
\	sysBP	diaBP		BMI	heartRate	glucose
count 423	38.000000	4238.000000	4219.00	00000 4	237.000000	3850.000000

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 89.875000 28.040000 83.000000 87.000000 max 295.000000 142.500000 56.800000 143.000000 394.000000 TenYearCHD count 4238.000000 mean 0.151958 std 0.359023 min 0.000000 50% 0.000000 50% 0.000000 max 1.000000 df male age education currentSmoker cigsPerDay BPMeds \ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0								
min 83.500000 48.000000 15.540000 44.000000 40.00000 25% 117.000000 75.000000 23.070000 68.000000 71.000000 50% 128.000000 82.000000 25.400000 75.000000 78.000000 max 295.000000 142.500000 56.800000 143.000000 394.00000 TenYearCHD count 4238.000000 mean 0.151958 std 0.359023 min 0.000000 25% 0.000000 75% 0.000000 max 1.000000 df male age education currentSmoker cigsPerDay BPMeds \ 0.000000 75% 0.000000 max 1.000000 df male age education currentSmoker cigsPerDay 0.00 1 39 4.0 0 0.0 0.0 0.0 2 1 488 1.0 1 20.0 0.0 2 1 488 1.0 1 20.0 0.0 3 0 61 3.0 1 30.0 0.0 4 0 46 3.0 1 20.0 0.0 4 0 46 3.0 1 20.0 0.0 4 0 46 3.0 1 20.0 0.0 4 1 51 3.0 1 43.0 0.0 4 234 1 51 3.0 1 43.0 0.0 4 235 0 48 2.0 1 20.0 NaN 4235 0 48 2.0 1 20.0 NaN 4236 0 44 1.0 1 1 15.0 0.0 4237 0 52 2.0 0 0 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 0 245.0 127.5 80.0 25.34 3 0 0 1 0 225.0 150.0 95.0	mean	132.35240	82.893	3464 2	5.802008	75.8789	24 8	1.966753
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 89.875000 28.040000 83.000000 87.000000 max 295.000000 142.500000 56.800000 143.000000 394.000000 TenYearCHD count 4238.0000000 mean 0.151958 std 0.359023 min 0.0000000 50% 0.0000000 75% 0.0000000 max 1.0000000 df male age education currentSmoker cigsPerDay BPMeds \ 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 3 0 61 3.0 1 30.0 0.0 4 0 46 3.0 1 30.0 0.0 4 0 46 3.0 1 23.0 0.0 4 0 46 3.0 1 23.0 0.0 4 0 46 3.0 1 23.0 0.0 4 0 46 3.0 1 23.0 0.0 4 0 46 3.0 1 23.0 0.0 4 0 46 3.0 1 50.0 0.0 4234 1 51 3.0 1 43.0 0.0 4235 0 48 2.0 1 20.0 NaN 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 0.0 4237 0 52 2.0 0 0 0 195.0 106.0 70.0 28.73 2 0 0 0 0 250.0 121.0 81.0 28.73 2 0 0 0 0 245.0 127.5 80.0 28.58	std	22.03809	7 11.910	0850	4.080111	12.0265	96 2	3.959998
TenYearCHD	min	83.50000	00 48.000	0000 1	5.540000	44.0000	00 4	0.00000
75% 144.00000 89.875000 28.040000 83.000000 87.00000 max 295.000000 142.500000 56.800000 143.000000 394.000000 TenYearCHD count 4238.000000 mean 0.151958 std 0.359023 min 0.000000 50% 0.000000 max 1.000000 df male age education currentSmoker cigsPerDay BPMeds \ 0 1 39 4.0 0 0 0.0 0.0 2 1 48 1.0 1 20.0 0.0 3 0 61 3.0 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 1 4233 1 50 1.0 1 23.0 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 0.0 prevalentStroke prevalentHyp diabetes totChol sysBP diaBP BMI \ 0 0 0 0 195.0 106.0 70.0 26.97 1 0 0 0 0 250.0 121.0 81.0 28.73 2 0 0 0 0 245.0 127.5 80.0 25.34 3 0 1 0 225.0 150.0 95.0	25%	117.00000	75.000	0000 2	3.070000	68.0000	00 7	1.000000
TenYearCHD count 4238.000000 mean 0.151958 std 0.359023 min 0.000000 50% 0.000000 75% 0.000000 max 1.000000 df male age education currentSmoker cigsPerDay BPMeds \ 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 3 0 61 3.0 1 30.0 0.0 4 0 46 3.0 1 23.0 0.0 4 0 46 3.0 1 23.0 0.0 4 0 46 3.0 1 23.0 0.0 4 0 46 3.0 1 23.0 0.0 4 0 46 3.0 1 23.0 0.0 4 233 1 50 1.0 1 20.0 0.0 4233 1 51 30 1.0 1 43.0 0.0 4235 0 48 2.0 1 43.0 0.0 4235 0 48 2.0 1 43.0 0.0 4236 0 44 1.0 1 15.0 0.0 4237 0 52 2.0 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 0 250.0 121.0 81.0 28.73 2 0 0 0 0 225.0 150.0 95.0	50%	128.00000	00 82.000	0000 2	5.400000	75.0000	00 7	8.000000
TenYearCHD count	75%	144.00000	00 89.875	5000 2	8.040000	83.0000	00 8	7.000000
count 4238.000000 mean 0.151958 std 0.359023 min 0.000000 0.000000 0.000000 50% 0.000000 0.000000 0.00 0.00 max 1.000000 0 0 0.0 0.0 max 1.000000 0 0.0 0.0 0.0 make age education currentSmoker cigsPerDay BPMeds \ 0 1 39 4.0 0 0.0 0.0 0.0 1 0 46 2.0 0 0.0 0.0 0.0 2 1 48 1.0 1 20.0 0.0 4 0 46 3.0 1 23.0 0.0 4 0 46 3.0 1 1.0 0.0 4233 1 50 1.0 1 1.0 0.0 4235 0 48 2.0 1 20.0 NaN 4236 0 <td>max</td> <td>295.00000</td> <td>00 142.500</td> <td>0000 5</td> <td>6.800000</td> <td>143.0000</td> <td>00 39</td> <td>94.000000</td>	max	295.00000	00 142.500	0000 5	6.800000	143.0000	00 39	94.000000
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 0 195.0 106.0 70.0 26.97 1	mean std min 25% 50% 75% max	4238.00000 0.15195 0.35902 0.00000 0.00000 0.00000	0 68 3 0 0 0 0 0 0					
0 1 39 4.0 0 0.0 0.0 0.0 1.0 1 0.0 1	df							
BMI \ 0	1 2 3 4 4233 4234 4235 4236 4237	1 39 0 46 1 48 0 61 0 46 1 50 1 51 0 48 0 44 0 52	4.0 2.0 1.0 3.0 3.0 1.0 3.0 2.0 1.0		0 0 1 1 1 1 1 1 1	0.0 0.0 20.0 30.0 23.0 1.0 43.0 20.0 15.0 0.0	0.0 0.0 0.0 0.0 0.0 0.0 NaN 0.0	
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			roke prev	alentHyp	diabetes	totChol	sysBP	diaBP
1 0 0 0 250.0 121.0 81.0 28.73 0 0 0 245.0 127.5 80.0 25.34 0 1 0 225.0 150.0 95.0 28.58	0	,	0	0	0	195.0	106.0	70.0
2 0 0 0 245.0 127.5 80.0 25.34 0 1 0 225.0 150.0 95.0 28.58	1		0	0	0	250.0	121.0	81.0
3 0 1 0 225.0 150.0 95.0 28.58	2		0	0	0	245.0	127.5	80.0
	3		0	1	0	225.0	150.0	95.0
			0	0	0	285.0	130.0	84.0

23.10						
• • •		• • •		• • •	• • •	
4233	0	1	0	313.0	179.0	92.0
25.97	U	1	U	313.0	179.0	92.0
4234	0	0	0	207.0	126.5	80.0
19.71						
4235	0	0	0	248.0	131.0	72.0
22.00	•		•	0.1.0	100 =	0 = 0
4236	0	0	0	210.0	126.5	87.0
19.16 4237	0	0	0	260 0	122 6	02 0
21.47	0	0	U	269.0	133.5	83.0

	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]

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	

df.isnull()

	nrewalent9	Stroke	prevalentHyp	diahotos	totChol	sysBP	diaBP
BMI	_ -	CIONC	prevarentilyp	arabetes	COCCIIOI	SYSDI	arabi
0		False	False	False	False	False	False
False		False	False	False	False	False	False
False		raise	raise	raise	raise	raise	raise
2		False	False	False	False	False	False
False		False	Ealaa	False	False	False	Folgo
False		raise	False	raise	raise	raise	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							
	1 + D - + -		- M				
0	heartRate False	glucos Fals					
1	False	Fals					
2	False	Fals					
3 4	False False	Fals Fals					
• • •							
4233	False	False					
4234	False	False					
4235 4236	False False	False Tru					
4237	False	Fals					

```
[4238 rows x 16 columns]
df.isnull().any()
male
                    False
                    False
age
                    True
education
currentSmoker
                    False
                     True
cigsPerDay
                     True
BPMeds
                    False
prevalentStroke
                    False
prevalentHyp
                    False
diabetes
                     True
totChol
                    False
sysBP
                    False
diaBP
                     True
BMI
                     True
heartRate
                     True
glucose
                    False
TenYearCHD
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()
                   0
male
                   0
age
                   0
education
currentSmoker
                   0
cigsPerDay
                   0
                   0
BPMeds
prevalentStroke
                   0
prevalentHyp
                   0
                   0
diabetes
                   0
totChol
                   0
sysBP
                   0
diaBP
                   0
BMI
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
            age education currentSmoker
                                             cigsPerDay BPMeds \
0
             39
         1
                        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
                                                         0.00000
                        1.0
                                                   1.0
4234
                        3.0
                                                         0.00000
             51
                                          1
                                                   43.0
         1
                                          1
4235
         0
             48
                        2.0
                                                   20.0 0.02963
4236
         0
             44
                                          1
                                                   15.0
                                                         0.00000
                        1.0
             52
4237
                        2.0
                                                    0.0
                                                         0.00000
     prevalentStroke
                        prevalentHyp diabetes totChol
                                                                 diaBP
                                                         sysBP
BMI \
                                                   195.0
                                                         106.0
                                                                   70.0
0
                     0
                                              0
26.97
                                                   250.0
                                                          121.0 81.0
28.73
2
                                                   245.0
                                                          127.5
                                                                  80.0
25.34
                                                          150.0
3
                                                   225.0
                                                                 95.0
28.58
                                                   285.0
                                                          130.0
                                                                   84.0
23.10
4233
                                                   313.0
                                                          179.0
                                                                   92.0
25.97
4234
                                                   207.0
                                                          126.5
                                                                   80.0
19.71
4235
                                                   248.0
                                                         131.0 72.0
22.00
4236
                                                   210.0
                                                         126.5 87.0
19.16
4237
                                                   269.0 133.5 83.0
21.47
      heartRate
                     glucose
0
           80.0
                   77.000000
```

```
1
           95.0
                  76.000000
2
                70.000000
           75.0
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)
y = df['TenYearCHD']
            age education currentSmoker
                                            cigsPerDay
                                                        BPMeds \
      male
0
         1
            39
                       4.0
                                         0
                                                   0.0
                                                         0.00000
1
         0
                        2.0
                                         0
                                                         0.00000
             46
                                                   0.0
2
            48
                                         1
         1
                       1.0
                                                  20.0
                                                         0.00000
3
         0
             61
                        3.0
                                         1
                                                  30.0
                                                         0.00000
4
         0
             46
                                         1
                                                  23.0 0.00000
                       3.0
            . . .
                        . . .
                                                   . . .
             50
                                                        0.00000
4233
         1
                       1.0
                                         1
                                                   1.0
                                         1
4234
         1
             51
                        3.0
                                                  43.0 0.00000
4235
                        2.0
                                         1
         0
             48
                                                  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
                                                  195.0
                                                         106.0 70.0
26.97
1
                                                  250.0
                                                         121.0 81.0
28.73
                                                  245.0
                                                         127.5
                                                                  80.0
25.34
3
                                                  225.0
                                                         150.0
                                                                  95.0
28.58
                                             0
                                                  285.0
                                                         130.0
                                                                  84.0
23.10
. . .
. . .
                                                  313.0 179.0 92.0
4233
25.97
4234
                                             0 207.0 126.5 80.0
19.71
```

```
4235
                    0
                                                  248.0 131.0
                                                                 72.0
22.00
4236
                                                  210.0 126.5 87.0
19.16
4237
                                                  269.0 133.5 83.0
21.47
      heartRate
                    glucose
0
           80.0
                77.000000
1
                  76.000000
           95.0
2
           75.0
                 70.000000
3
           65.0 103.000000
4
           85.0 85.000000
            . . .
. . .
           66.0 86.000000
4233
           65.0 68.000000
4234
           84.0
                86.000000
4235
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
4089
        0
3444
        0
466
        0
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
```