

KNN K Nearest Neighbour

```
In [2]: #Name: Yash Pravin Gadbail  
#Roll no. : 35  
#Sec: 3rd A  
#Sub : ET 1  
#Date:05/10/2024
```

```
In [4]: #Aim: To Perform Operation on KNN K Nearest Neighbour
```

```
In [6]: 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')
```

```
In [8]: df=pd.read_csv("C:\\Users\\OneDrive\\Desktop\\framingham.csv")
```

```
In [10]: df.head()
```

```
Out[10]:
```

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp
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	0	0
3	0	61	3.0	1	30.0	0.0	0	1
4	0	46	3.0	1	23.0	0.0	0	0



In [12]: df.describe

```
Out[12]: <bound method NDFrame.describe of
r  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  BM
I  \
0      0      0      0      195.0  106.0  70.0  26.9
7
1      0      0      0      250.0  121.0  81.0  28.7
3
2      0      0      0      245.0  127.5  80.0  25.3
4
3      0      1      0      225.0  150.0  95.0  28.5
8
4      0      0      0      285.0  130.0  84.0  23.1
0
...    ...    ...    ...    ...    ...    ...
...
4235   0      0      0      248.0  131.0  72.0  22.0
0
4236   0      0      0      210.0  126.5  87.0  19.1
6
4237   0      0      0      269.0  133.5  83.0  21.4
7
4238   0      1      0      185.0  141.0  98.0  25.6
0
4239   0      0      0      196.0  133.0  86.0  20.9
1

      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]>
```

```
In [14]: 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
```

```
In [16]: df.isna().sum()
```

```
Out[16]: 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
```

In [18]: df

Out[18]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalent
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	
...
4235	0	48	2.0	1	20.0	NaN	0	
4236	0	44	1.0	1	15.0	0.0	0	
4237	0	52	2.0	0	0.0	0.0	0	
4238	1	40	3.0	0	0.0	0.0	0	
4239	0	39	3.0	1	30.0	0.0	0	

4240 rows × 16 columns



Missing Value Treatment

In [75]: df['education'].fillna(value = df['education'].mean(),inplace=True)

In [87]: df['glucose'].fillna(value = df['glucose'].mean(),inplace=True)

In [99]: df['heartRate'].fillna(value = df['heartRate'].mean(),inplace=True)

In [101]: df['BMI'].fillna(value = df['BMI'].mean(),inplace=True)

In [103]: df['cigsPerDay'].fillna(value = df['cigsPerDay'].mean(),inplace=True)

In [105]: df['totChol'].fillna(value = df['totChol'].mean(),inplace=True)

In [107]: df['BPMeds'].fillna(value = df['BPMeds'].mean(),inplace=True)

```
In [109]: df.isna().sum()
```

```
Out[109]: 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
```

```
In [115]: x = df.drop("TenYearCHD",axis=1)
y = df['TenYearCHD']
```

```
In [117]: x
```

```
Out[117]:
```

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

4240 rows × 15 columns



Train Test Split

```
In [120]: x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_s
```

In [122]: y_train

```
Out[122]: 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
```

KNN Classifier

```
In [125]: 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

In []: