KNN k nearest

In [2]:
```
#Aim : To perform and find the accuracy of K-Nearest Neighbors Algorithm i.e. KNN C
```

In [4]:
```
#Name: Leena Rajeshwar Kale
#Roll No.:71
#Sec: C
#Subject:ET - 1
```

In [1]:
```python
import pandas as pd
```

In [2]:
```python
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 [3]:
```python
import os
```

In [4]:
```python
os.getcwd()
```

Out[4]:
```
'C:\\Users\\dishi\\Downloads'
```

In [5]:
```python
os.chdir("C:\\Users\\dishi\\Downloads")
```

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

In [20]:
```python
df.tail()
```

Out[20]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp
4233	1	50	1.0	1	1.0	0.0	0	1
4234	1	51	3.0	1	43.0	0.0	0	0
4235	0	48	2.0	1	20.0	NaN	0	0
4236	0	44	1.0	1	15.0	0.0	0	0
4237	0	52	2.0	0	0.0	0.0	0	0

In [18]:
```python
df.head()
```

Out[18]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	dia
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 [22]: `df.isna().sum()`

Out[22]:
```
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 [24]: `df.isnull()`

Out[24]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...
4233	False	False	False	False	False	False	False	False
4234	False	False	False	False	False	False	False	False
4235	False	False	False	False	False	True	False	False
4236	False	False	False	False	False	False	False	False
4237	False	False	False	False	False	False	False	False

4238 rows × 16 columns

Missing Value Tratment

```python
In [29]: df['glucose'].fillna(value = df['glucose'].mean(),inplace=True)
```

```python
In [31]: df['education'].fillna(value = df['education'].mean(),inplace=True)
```

```python
In [33]: df['heartRate'].fillna(value = df['heartRate'].mean(),inplace=True)
```

```python
In [35]: df['BMI'].fillna(value = df['BMI'].mean(),inplace=True)
```

```python
In [37]: df['cigsPerDay'].fillna(value = df['cigsPerDay'].mean(),inplace=True)
```

```python
In [39]: df['totChol'].fillna(value = df['totChol'].mean(),inplace=True)
```

```python
In [41]: df['BPMeds'].fillna(value = df['BPMeds'].mean(),inplace=True)
```

```python
In [43]: df.isna().sum()
```

```
Out[43]: 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
```

```python
In [45]: #Splitting the dependent and independent variables.
         x = df.drop("TenYearCHD",axis=1)
         y = df['TenYearCHD']
```

```python
In [47]: y
```

```
Out[47]: 0       0
         1       0
         2       0
         3       1
         4       0
                ..
         4233    1
         4234    0
         4235    0
         4236    0
         4237    0
         Name: TenYearCHD, Length: 4238, dtype: int64
```

Train Test Split

```python
In [50]: x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=42)
```

```python
In [52]: x_train
```

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp
3252	1	40	4.0	1	30.0	0.0	0	0
3946	0	57	2.0	0	0.0	0.0	0	1
1261	0	47	1.0	0	0.0	0.0	0	0
2536	1	41	2.0	1	30.0	0.0	0	0
4089	0	64	1.0	0	0.0	0.0	0	1
...	...	...	...	...	...	...	...	...
3444	0	36	1.0	1	5.0	0.0	0	1
466	0	57	3.0	1	15.0	0.0	0	0
3092	0	60	2.0	0	0.0	0.0	0	1
3772	1	39	2.0	1	10.0	0.0	0	0
860	0	35	2.0	0	0.0	0.0	0	0

3390 rows × 15 columns

In [54]:
```python
y_train
```

Out[54]:
```
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
```

KNN Classifier

In [57]:
```python
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
```

In [ ]: