In [50]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [51]:
```
from sklearn.linear_model import LogisticRegression
```

In [52]:
```
df=pd.read_csv(r"C:\Users\Admin\Downloads\C4_framingham - C4_framingham.csv")
df
```

Out[52]:

| | 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 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... |
| **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 |

4238 rows × 16 columns

In [53]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4238 entries, 0 to 4237
Data columns (total 16 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   male             4238 non-null   int64  
 1   age              4238 non-null   int64  
 2   education        4133 non-null   float64
 3   currentSmoker    4238 non-null   int64  
 4   cigsPerDay       4209 non-null   float64
 5   BPMeds           4185 non-null   float64
 6   prevalentStroke  4238 non-null   int64  
 7   prevalentHyp     4238 non-null   int64  
 8   diabetes         4238 non-null   int64  
 9   totChol          4188 non-null   float64
 10  sysBP            4238 non-null   float64
 11  diaBP            4238 non-null   float64
 12  BMI              4219 non-null   float64
 13  heartRate        4237 non-null   float64
 14  glucose          3850 non-null   float64
 15  TenYearCHD       4238 non-null   int64  
dtypes: float64(9), int64(7)
memory usage: 529.9 KB
```

In [54]:
```
df1=df.dropna()
```

In [55]:
```
d1=df1.iloc[:,0:5]
d2=df1.iloc[:,-1]
```

In [56]:
```
d1.shape
```

Out[56]: (3656, 5)

In [57]:
```
d2.shape
```

Out[57]: (3656,)

In [58]:
```
from sklearn.preprocessing import StandardScaler
```

In [59]:
```
a=StandardScaler().fit_transform(d1)
```

In [60]:
```
lr=LogisticRegression()
lr.fit(a,d2)
```

Out[60]: LogisticRegression()

In [61]:
```
obs=[[10,12,13,14,5]]
```

In [62]:
```
pdt=lr.predict(obs)
print(pdt)
```

```
[1]
```

In [63]:
```
lr.classes_
```

Out[63]:
```
array([0, 1], dtype=int64)
```

In [64]:
```
lr.predict_proba(obs)[0][0]
```

Out[64]:
```
0.0001619639119773142
```

In [65]:
```
lr.predict_proba(obs)[0][1]
```

Out[65]:
```
0.9998380360880227
```

In [ ]:

In [ ]: