

```
In [1]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [2]: from sklearn.linear_model import LogisticRegression
```

```
In [3]: df=pd.read_csv("C4_framingham.csv")
df
```

```
Out[3]:
```

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabe
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 [4]: df=df.dropna()
df
```

```
Out[4]:
```

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabe
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	
...
4231	1	58	3.0	0	0.0	0.0	0	1	
4232	1	68	1.0	0	0.0	0.0	0	1	

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabe
4233	1	50	1.0	1	1.0	0.0	0	1	
4234	1	51	3.0	1	43.0	0.0	0	0	
4237	0	52	2.0	0	0.0	0.0	0	0	

3656 rows × 16 columns

In [5]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3656 entries, 0 to 4237
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   male                  3656 non-null   int64
1   age                   3656 non-null   int64
2   education             3656 non-null   float64
3   currentSmoker         3656 non-null   int64
4   cigsPerDay            3656 non-null   float64
5   BPMeds                3656 non-null   float64
6   prevalentStroke       3656 non-null   int64
7   prevalentHyp          3656 non-null   int64
8   diabetes              3656 non-null   int64
9   totChol               3656 non-null   float64
10  sysBP                 3656 non-null   float64
11  diaBP                 3656 non-null   float64
12  BMI                   3656 non-null   float64
13  heartRate             3656 non-null   float64
14  glucose               3656 non-null   float64
15  TenYearCHD            3656 non-null   int64
dtypes: float64(9), int64(7)
memory usage: 485.6 KB
```

In [6]:

```
df.columns
```

```
Out[6]: Index(['male', 'age', 'education', 'currentSmoker', 'cigsPerDay', 'BPMeds',
              'prevalentStroke', 'prevalentHyp', 'diabetes', 'totChol', 'sysBP',
              'diaBP', 'BMI', 'heartRate', 'glucose', 'TenYearCHD'],
              dtype='object')
```

In [7]:

```
feature_matrix=df[['male', 'age', 'education', 'currentSmoker', 'cigsPerDay', 'BPMeds',
                  'prevalentStroke', 'prevalentHyp', 'diabetes', 'totChol', 'sysBP',
                  'diaBP', 'BMI', 'heartRate', 'glucose']]
target_vector=df['TenYearCHD']
```

In [8]:

```
feature_matrix.shape
```

```
Out[8]: (3656, 15)
```

In [9]:

```
target_vector.shape
```

Out[9]: (3656,)

```
In [10]: from sklearn.preprocessing import StandardScaler
```

```
In [11]: fs=StandardScaler().fit_transform(feature_matrix)
```

```
In [12]: logr=LogisticRegression()  
logr.fit(fs,target_vector)
```

Out[12]: LogisticRegression()

```
In [13]: observation=[[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15]]
```

```
In [14]: prediction=logr.predict(observation)  
print(prediction)
```

[1]

```
In [15]: logr.classes_
```

Out[15]: array([0, 1], dtype=int64)

```
In [16]: logr.predict_proba(observation)[0][0]
```

Out[16]: 0.0002214783507201723

```
In [17]: logr.predict_proba(observation)
```

Out[17]: array([[2.21478351e-04, 9.99778522e-01]])