LOGISTIC REGRESSION

```
In [2]:  #Name : Shreya Sharma
         #Roll no. : 46
         #Sectin : 3B
         #Date : 05/10/2024
```

```
In [3]:  #Aim : To perform operation on logestic regression
```

```
In [3]:  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 [4]:  import os
```

```
In [5]:  os.getcwd()
```

```
Out[5]:  'C:\\Users\\pravi'
```

```
In [6]:  os.chdir("C:\\Users\\pravi\\Desktop")
```

```
In [11]: df=pd.read_csv("framingham.csv")
```

```
In [13]: df.head()
```

Out[13]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabetes	totChol	sysBP	diaBP	
0	1	39	4.0	0	0.0	0.0	0	0	0	195.0	106.0	70.0	26
1	0	46	2.0	0	0.0	0.0	0	0	0	250.0	121.0	81.0	28
2	1	48	1.0	1	20.0	0.0	0	0	0	245.0	127.5	80.0	25
3	0	61	3.0	1	30.0	0.0	0	1	0	225.0	150.0	95.0	28
4	0	46	3.0	1	23.0	0.0	0	0	0	285.0	130.0	84.0	23

```
In [15]: 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 [17]: df.isna().sum()
```

```
Out[17]:  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 [19]: `df`

Out[19]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabetes	totChol	sysBP	diaBP
0	1	39	4.0	0	0.0	0.0	0	0	0	195.0	106.0	70.0
1	0	46	2.0	0	0.0	0.0	0	0	0	250.0	121.0	81.0
2	1	48	1.0	1	20.0	0.0	0	0	0	245.0	127.5	80.0
3	0	61	3.0	1	30.0	0.0	0	1	0	225.0	150.0	95.0
4	0	46	3.0	1	23.0	0.0	0	0	0	285.0	130.0	84.0
...	...	...	...	...	...	...	...	...	...	...	...	...
4233	1	50	1.0	1	1.0	0.0	0	1	0	313.0	179.0	92.0
4234	1	51	3.0	1	43.0	0.0	0	0	0	207.0	126.5	80.0
4235	0	48	2.0	1	20.0	NaN	0	0	0	248.0	131.0	72.0
4236	0	44	1.0	1	15.0	0.0	0	0	0	210.0	126.5	87.0
4237	0	52	2.0	0	0.0	0.0	0	0	0	269.0	133.5	83.0

4238 rows × 16 columns
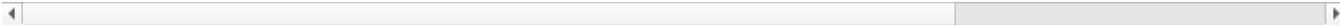
In [21]: `df['glucose'].fillna(value = df['glucose'].mean(),inplace=True)`

In [23]: `df['education'].fillna(value = df['education'].mean(),inplace=True)`

In [25]: `df['heartRate'].fillna(value = df['heartRate'].mean(),inplace=True)`

In [27]: `df['BMI'].fillna(value = df['BMI'].mean(),inplace=True)`

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

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

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

In [35]: `df.isna().sum()`

```
Out[35]:  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 [37]:
```
x = df.drop("TenYearCHD",axis=1)
y = df['TenYearCHD']
```

```
In [39]:  x
```

Out[39]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabetes	totChol	sysBP	diaBP
0	1	39	4.0	0	0.0	0.00000	0	0	0	195.0	106.0	70.0
1	0	46	2.0	0	0.0	0.00000	0	0	0	250.0	121.0	81.0
2	1	48	1.0	1	20.0	0.00000	0	0	0	245.0	127.5	80.0
3	0	61	3.0	1	30.0	0.00000	0	1	0	225.0	150.0	95.0
4	0	46	3.0	1	23.0	0.00000	0	0	0	285.0	130.0	84.0
...	...	...	...	...	...	...	...	...	...	...	...	...
4233	1	50	1.0	1	1.0	0.00000	0	1	0	313.0	179.0	92.0
4234	1	51	3.0	1	43.0	0.00000	0	0	0	207.0	126.5	80.0
4235	0	48	2.0	1	20.0	0.02963	0	0	0	248.0	131.0	72.0
4236	0	44	1.0	1	15.0	0.00000	0	0	0	210.0	126.5	87.0
4237	0	52	2.0	0	0.0	0.00000	0	0	0	269.0	133.5	83.0

4238 rows × 15 columns

```
In [41]:  x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=42)
```

```
In [43]:  x_train
```

Out[43]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabetes	totChol	sysBP	diaBP
3252	1	40	4.0	1	30.0	0.0	0	0	0	205.0	131.0	81.0
3946	0	57	2.0	0	0.0	0.0	0	1	0	250.0	152.5	92.5
1261	0	47	1.0	0	0.0	0.0	0	0	0	230.0	123.0	71.0
2536	1	41	2.0	1	30.0	0.0	0	0	0	228.0	113.0	82.5
4089	0	64	1.0	0	0.0	0.0	0	1	0	232.0	149.5	84.0
...	...	...	...	...	...	...	...	...	...	...	...	...
3444	0	36	1.0	1	5.0	0.0	0	1	0	222.0	147.0	94.0
466	0	57	3.0	1	15.0	0.0	0	0	0	250.0	125.0	74.0
3092	0	60	2.0	0	0.0	0.0	0	1	0	298.0	133.0	89.0
3772	1	39	2.0	1	10.0	0.0	0	0	0	215.0	102.0	64.5
860	0	35	2.0	0	0.0	0.0	0	0	0	248.0	107.0	73.0

3390 rows × 15 columns

```
In [45]:  y_train
```

```
Out[45]:  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
```

```
In [47]:  from sklearn.linear_model import LogisticRegression
          model = LogisticRegression().fit(x_train,y_train)
          model.score(x_train, y_train)
```

```
Out[47]:  0.8495575221238938
```

```
In [ ]:
```