

```
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	0
1	0	46	2.0	0	0.0	0.0	0	0	0
2	1	48	1.0	1	20.0	0.0	0	0	0
3	0	61	3.0	1	30.0	0.0	0	1	1
4	0	46	3.0	1	23.0	0.0	0	0	0
...
4233	1	50	1.0	1	1.0	0.0	0	1	1
4234	1	51	3.0	1	43.0	0.0	0	0	0
4235	0	48	2.0	1	20.0	NaN	0	0	0
4236	0	44	1.0	1	15.0	0.0	0	0	0
4237	0	52	2.0	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	0
1	0	46	2.0	0	0.0	0.0	0	0	0
2	1	48	1.0	1	20.0	0.0	0	0	0
3	0	61	3.0	1	30.0	0.0	0	1	1
4	0	46	3.0	1	23.0	0.0	0	0	0
...
4231	1	58	3.0	0	0.0	0.0	0	1	1
4232	1	68	1.0	0	0.0	0.0	0	1	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]])
```

```
In [18]: df['TenYearCHD'].value_counts()
```

```
Out[18]: 0    3099
1     557
Name: TenYearCHD, dtype: int64
```

```
In [19]: x=df[['male', 'age', 'education', 'currentSmoker', 'cigsPerDay', 'BPMeds',
           'prevalentStroke', 'prevalentHyp', 'diabetes', 'totChol', 'sysBP',
           'diaBP', 'BMI', 'heartRate', 'glucose']]
y=df['TenYearCHD']
```

```
In [20]: from sklearn.model_selection import train_test_split
```

```
In [21]: x_train,x_test,y_train,y_test=train_test_split(x,y,train_size=0.70)
```

```
In [22]: from sklearn.ensemble import RandomForestClassifier
```

```
In [23]: rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

```
Out[23]: RandomForestClassifier()
```

```
In [24]: parameters={'max_depth':[1,2,3,4,5],
                 'min_samples_leaf':[5,10,15,20,25],
                 'n_estimators':[10,20,30,40,50]
                }
```

```
In [25]: from sklearn.model_selection import GridSearchCV
grid_search =GridSearchCV(estimator=rfc,param_grid=parameters, cv=2,scoring="accuracy")
grid_search.fit(x_train,y_train)
```

```
Out[25]: GridSearchCV(cv=2, estimator=RandomForestClassifier(),
                      param_grid={'max_depth': [1, 2, 3, 4, 5],
                                  'min_samples_leaf': [5, 10, 15, 20, 25],
                                  'n_estimators': [10, 20, 30, 40, 50]},
                      scoring='accuracy')
```

```
In [26]: grid_search.best_score_
```

```
Out[26]: 0.8483785550234558
```

```
In [27]: rfc_best=grid_search.best_estimator_
```

```
In [28]: from sklearn.tree import plot_tree
```

```
plt.figure(figsize=(80,40))
plot_tree(rfc_best.estimators_[5],feature_names=x.columns,class_names=[ 'Yes' , 'No' ],fill
```

```
Out[28]: [Text(2332.44, 1993.2, 'age <= 48.5\n gini = 0.259\n samples = 1609\n value = [2168, 391]\n class = Yes'),
          Text(1227.6, 1630.800000000002, 'totChol <= 304.5\n gini = 0.132\n samples = 785\n value = [1211, 93]\n class = Yes'),
          Text(714.24, 1268.4, 'currentSmoker <= 0.5\n gini = 0.115\n samples = 750\n value = [1169, 76]\n class = Yes'),
          Text(357.12, 906.0, 'age <= 41.5\n gini = 0.086\n samples = 335\n value = [527, 25]\n class = Yes'),
          Text(178.56, 543.599999999999, 'prevalentHyp <= 0.5\n gini = 0.04\n samples = 144\n value = [237, 5]\n class = Yes'),
          Text(89.28, 181.199999999982, 'gini = 0.046\n samples = 124\n value = [205, 5]\n class = Yes'),
          Text(267.840000000003, 181.199999999982, 'gini = 0.0\n samples = 20\n value = [32, 0]\n class = Yes'),
          Text(535.680000000001, 543.599999999999, 'male <= 0.5\n gini = 0.121\n samples = 191\n value = [290, 20]\n class = Yes'),
          Text(446.4, 181.199999999982, 'gini = 0.164\n samples = 121\n value = [182, 18]\n class = Yes'),
          Text(624.96, 181.199999999982, 'gini = 0.036\n samples = 70\n value = [108, 2]\n class =
```

```

Yes'),
Text(1071.3600000000001, 906.0, 'totChol <= 284.5\ngini = 0.136\nsamples = 415\nvalue =
[642, 51]\nclass = Yes'),
Text(892.8, 543.5999999999999, 'age <= 47.5\ngini = 0.123\nsamples = 386\nvalue = [595,
42]\nclass = Yes'),
Text(803.52, 181.1999999999982, 'gini = 0.109\nsamples = 365\nvalue = [571, 35]\nclass
= Yes'),
Text(982.08, 181.1999999999982, 'gini = 0.35\nsamples = 21\nvalue = [24, 7]\nclass = Y
es'),
Text(1249.92, 543.5999999999999, 'cigsPerDay <= 17.5\ngini = 0.27\nsamples = 29\nvalue
= [47, 9]\nclass = Yes'),
Text(1160.64, 181.1999999999982, 'gini = 0.0\nsamples = 9\nvalue = [19, 0]\nclass = Ye
s'),
Text(1339.2, 181.1999999999982, 'gini = 0.368\nsamples = 20\nvalue = [28, 9]\nclass =
Yes'),
Text(1740.96, 1268.4, 'age <= 41.5\ngini = 0.41\nsamples = 35\nvalue = [42, 17]\nclass
= Yes'),
Text(1517.76, 906.0, 'sysBP <= 127.0\ngini = 0.26\nsamples = 13\nvalue = [22, 4]\nclass
= Yes'),
Text(1428.48, 543.5999999999999, 'gini = 0.208\nsamples = 8\nvalue = [15, 2]\nclass = Y
es'),
Text(1607.04, 543.5999999999999, 'gini = 0.346\nsamples = 5\nvalue = [7, 2]\nclass = Ye
s'),
Text(1964.16, 906.0, 'heartRate <= 73.5\ngini = 0.478\nsamples = 22\nvalue = [20, 13]\n
class = Yes'),
Text(1785.6, 543.5999999999999, 'diaBP <= 78.0\ngini = 0.498\nsamples = 10\nvalue = [7,
8]\nclass = No'),
Text(1696.32, 181.1999999999982, 'gini = 0.49\nsamples = 5\nvalue = [4, 3]\nclass = Ye
s'),
Text(1874.88, 181.1999999999982, 'gini = 0.469\nsamples = 5\nvalue = [3, 5]\nclass = N
o'),
Text(2142.720000000003, 543.5999999999999, 'BMI <= 26.475\ngini = 0.401\nsamples = 12
\nvalue = [13, 5]\nclass = Yes'),
Text(2053.44, 181.1999999999982, 'gini = 0.49\nsamples = 5\nvalue = [4, 3]\nclass = Ye
s'),
Text(2232.0, 181.1999999999982, 'gini = 0.298\nsamples = 7\nvalue = [9, 2]\nclass = Ye
s'),
Text(3437.28, 1630.800000000002, 'prevalentHyp <= 0.5\ngini = 0.362\nsamples = 824\nva
lue = [957, 298]\nclass = Yes'),
Text(2990.88, 1268.4, 'totChol <= 272.5\ngini = 0.294\nsamples = 478\nvalue = [605, 13
2]\nclass = Yes'),
Text(2678.4, 906.0, 'age <= 66.5\ngini = 0.318\nsamples = 372\nvalue = [456, 113]\nclas
s = Yes'),
Text(2499.84, 543.5999999999999, 'currentSmoker <= 0.5\ngini = 0.304\nsamples = 362\nva
lue = [449, 103]\nclass = Yes'),
Text(2410.56, 181.1999999999982, 'gini = 0.239\nsamples = 205\nvalue = [273, 44]\nclas
s = Yes'),
Text(2589.12, 181.1999999999982, 'gini = 0.376\nsamples = 157\nvalue = [176, 59]\nclas
s = Yes'),
Text(2856.96, 543.5999999999999, 'cigsPerDay <= 0.5\ngini = 0.484\nsamples = 10\nvalue
= [7, 10]\nclass = No'),
Text(2767.68, 181.1999999999982, 'gini = 0.42\nsamples = 5\nvalue = [3, 7]\nclass = N
o'),
Text(2946.240000000002, 181.1999999999982, 'gini = 0.49\nsamples = 5\nvalue = [4, 3]
\nclass = Yes'),
Text(3303.36, 906.0, 'sysBP <= 149.0\ngini = 0.201\nsamples = 106\nvalue = [149, 19]\n
class = Yes'),
Text(3214.08, 543.5999999999999, 'BMI <= 20.845\ngini = 0.181\nsamples = 100\nvalue =
[143, 16]\nclass = Yes'),
Text(3124.8, 181.1999999999982, 'gini = 0.48\nsamples = 5\nvalue = [2, 3]\nclass = N
o'),
Text(3303.36, 181.1999999999982, 'gini = 0.155\nsamples = 95\nvalue = [141, 13]\nclass
= Yes'),
Text(3392.64, 543.5999999999999, 'gini = 0.444\nsamples = 6\nvalue = [6, 3]\nclass = Ye
s'),

```

```

Text(3883.68, 1268.4, 'diaBP <= 72.75\ngini = 0.436\nsamples = 346\nvalue = [352, 166]
\nclass = Yes'),
Text(3660.48, 906.0, 'glucose <= 75.5\ngini = 0.278\nsamples = 11\nvalue = [3, 15]\ncla
ss = No'),
Text(3571.2, 543.5999999999999, 'gini = 0.444\nsamples = 5\nvalue = [3, 6]\nnclass = N
o'),
Text(3749.76, 543.5999999999999, 'gini = 0.0\nsamples = 6\nvalue = [0, 9]\nnclass = N
o'),
Text(4106.88, 906.0, 'BMI <= 28.865\ngini = 0.422\nsamples = 335\nvalue = [349, 151]\nnc
lass = Yes'),
Text(3928.32, 543.5999999999999, 'sysBP <= 145.25\ngini = 0.364\nsamples = 209\nvalue =
[235, 74]\nnclass = Yes'),
Text(3839.04, 181.1999999999982, 'gini = 0.199\nsamples = 65\nvalue = [87, 11]\nnclass
= Yes'),
Text(4017.6, 181.1999999999982, 'gini = 0.419\nsamples = 144\nvalue = [148, 63]\nnclass
= Yes'),
Text(4285.440000000005, 543.5999999999999, 'male <= 0.5\ngini = 0.481\nsamples = 126\nn
value = [114, 77]\nnclass = Yes'),
Text(4196.16, 181.1999999999982, 'gini = 0.446\nsamples = 88\nvalue = [87, 44]\nnclass
= Yes'),
Text(4374.72, 181.1999999999982, 'gini = 0.495\nsamples = 38\nvalue = [27, 33]\nnclass
= No')
]

```

