In [1]:
```
#Naive based lung cancer
```

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

In [49]:
```
df=pd.read_csv("survey lung cancer.csv")
```

In [50]:
```
df
```

Out[50]:

	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	CHRONIC DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL CONSUMING	COUGHING	SHOR OF B
0	M	69	1	2	2	1	1	2	1	2	2	2	
1	M	74	2	1	1	1	2	2	2	1	1	1	
2	F	59	1	1	1	2	1	2	1	2	1	2	
3	M	63	2	2	2	1	1	1	1	1	2	1	
4	F	63	1	2	1	1	1	1	1	2	1	2	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
304	F	56	1	1	1	2	2	2	1	1	2	2	
305	M	70	2	1	1	1	1	2	2	2	2	2	
306	M	58	2	1	1	1	1	1	2	2	2	2	
307	M	67	2	1	2	1	1	2	2	1	2	2	
308	M	62	1	1	1	2	1	2	2	2	2	1	

309 rows × 16 columns

In [51]:
```
df.shape
```

Out[51]:
```
(309, 16)
```

In [52]:
```
df
```

Out[52]:

	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	CHRONIC DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL CONSUMING	COUGHING	SHOR OF B
0	M	69	1	2	2	1	1	2	1	2	2	2	
1	M	74	2	1	1	1	2	2	2	1	1	1	
2	F	59	1	1	1	2	1	2	1	2	1	2	
3	M	63	2	2	2	1	1	1	1	1	2	1	
4	F	63	1	2	1	1	1	1	1	2	1	2	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
304	F	56	1	1	1	2	2	2	1	1	2	2	
305	M	70	2	1	1	1	1	2	2	2	2	2	
306	M	58	2	1	1	1	1	1	2	2	2	2	
307	M	67	2	1	2	1	1	2	2	1	2	2	
308	M	62	1	1	1	2	1	2	2	2	2	1	

309 rows × 16 columns

In [53]:
```
data=pd.DataFrame(df)
```

In [54]:
```
data
```

Out[54]:

	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	CHRONIC DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL CONSUMING	COUGHING	SHOR OF B
0	M	69	1	2	2	1	1	2	1	2	2	2	
1	M	74	2	1	1	1	2	2	2	1	1	1	
2	F	59	1	1	1	2	1	2	1	2	1	2	
3	M	63	2	2	2	1	1	1	1	1	2	1	
4	F	63	1	2	1	1	1	1	1	2	1	2	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
304	F	56	1	1	1	2	2	2	1	1	2	2	
305	M	70	2	1	1	1	1	2	2	2	2	2	
306	M	58	2	1	1	1	1	1	2	2	2	2	
307	M	67	2	1	2	1	1	2	2	1	2	2	
308	M	62	1	1	1	2	1	2	2	2	2	1	

309 rows × 16 columns

```
data
```

In [55]:
```
import numpy as np
```

In [56]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings # for avoid unwanted warnings
warnings.filterwarnings('ignore')
from IPython.core.display import display, HTML # for HTML tag use in python
```

In [57]:
```
from sklearn import preprocessing
encoder = preprocessing.LabelEncoder()
for i in data.columns:
    if isinstance(data[i][0], str):
         data[i] = encoder.fit_transform(data[i])
```

In [58]:
```
data
```

Out[58]:

	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	CHRONIC DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL CONSUMING	COUGHING	SHOR OF B
0	1	69	1	2	2	1	1	2	1	2	2	2	
1	1	74	2	1	1	1	2	2	2	1	1	1	
2	0	59	1	1	1	2	1	2	1	2	1	2	
3	1	63	2	2	2	1	1	1	1	1	2	1	
4	0	63	1	2	1	1	1	1	1	2	1	2	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
304	0	56	1	1	1	2	2	2	1	1	2	2	
305	1	70	2	1	1	1	1	2	2	2	2	2	
306	1	58	2	1	1	1	1	1	2	2	2	2	
307	1	67	2	1	2	1	1	2	2	1	2	2	
308	1	62	1	1	1	2	1	2	2	2	2	1	

309 rows × 16 columns

In [59]:
```
y = data.iloc[:,[-1]]
x = data.iloc[:,:-1]
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,y,random_state=3,test_size=0.2)
```

In [60]:
```
x_train
```

Out[60]:

	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	CHRONIC DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL CONSUMING	COUGHING	SHOR OF B
114	0	72	1	2	1	1	1	2	1	2	2	2	
265	0	60	2	2	2	2	1	2	2	1	1	1	
105	0	60	1	1	1	1	2	2	1	1	1	1	
224	0	62	2	1	1	2	1	2	2	2	2	2	
83	0	81	1	1	1	2	2	1	2	1	2	2	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
277	0	87	1	1	1	1	2	2	1	1	1	1	
256	1	60	2	1	1	1	1	2	2	2	2	2	
131	0	61	1	2	2	2	1	1	2	2	1	2	
249	1	68	2	1	2	1	1	2	1	1	1	1	
152	0	74	1	2	2	2	2	2	1	2	2	1	

247 rows × 15 columns

In [61]:
```
x_train.shape, y_train.shape, x_test.shape, y_test.shape
```

Out[61]: ((247, 15), (247, 1), (62, 15), (62, 1))

In [69]:
```
x.nunique()
```

Out[69]:
```
GENDER                    2
AGE                      39
SMOKING                   2
YELLOW_FINGERS            2
ANXIETY                   2
PEER_PRESSURE             2
CHRONIC DISEASE           2
FATIGUE                   2
ALLERGY                   2
WHEEZING                  2
ALCOHOL CONSUMING         2
COUGHING                  2
SHORTNESS OF BREATH       2
SWALLOWING DIFFICULTY     2
CHEST PAIN                2
dtype: int64
```

In [63]:
```
from sklearn.naive_bayes import GaussianNB
```

In [64]:
```
classifier = GaussianNB()
classifier.fit(x_train,y_train)
```

Out[64]: GaussianNB()

In [68]:
```
model_accuracy=round((classifier.score(x_test,y_test)*100),2)
HTML(f'Model Accuracy : <b><u>{ model_accuracy }%</b></u>')
```

Out[68]: Model Accuracy : **87.1%**

In [ ]:
```

```