

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
df=pd.read_csv('cancer.csv')
```

```
df.head()
```

	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	\
0	M	69	1	2	2	1	
1	M	74	2	1	1	1	
2	F	59	1	1	1	2	
3	M	63	2	2	2	1	
4	F	63	1	2	1	1	

	CHRONIC DISEASE COUGHING \	FATIGUE	ALLERGY	WHEEZING	ALCOHOL CONSUMING
0		1	2	1	2
2					
1		2	2	2	1
1					
2		1	2	1	2
2					
3		1	1	1	1
1					
4		1	1	1	2
2					

	SHORTNESS OF BREATH	SWALLOWING DIFFICULTY	CHEST PAIN	LUNG_CANCER	
0	2		2	2	YES
1	2		2	2	YES
2	2		1	2	NO
3	1		2	2	NO
4	2		1	1	NO

```
df.isnull().sum()
```

GENDER	0
AGE	0
SMOKING	0
YELLOW_FINGERS	0
ANXIETY	0
PEER_PRESSURE	0
CHRONIC DISEASE	0

```

FATIGUE          0
ALLERGY          0
WHEEZING         0
ALCOHOL CONSUMING 0
COUGHING         0
SHORTNESS OF BREATH 0
SWALLOWING DIFFICULTY 0
CHEST PAIN       0
LUNG_CANCER      0
dtype: int64

df.isna().sum()

GENDER          0
AGE             0
SMOKING         0
YELLOW_FINGERS  0
ANXIETY         0
PEER_PRESSURE   0
CHRONIC DISEASE 0
FATIGUE         0
ALLERGY         0
WHEEZING        0
ALCOHOL CONSUMING 0
COUGHING        0
SHORTNESS OF BREATH 0
SWALLOWING DIFFICULTY 0
CHEST PAIN      0
LUNG_CANCER     0
dtype: int64

df['GENDER']=df['GENDER'].map({'M':1,'F':2})
df['LUNG_CANCER']=df['LUNG_CANCER'].map({'YES':1,'NO':0})

from sklearn.model_selection import train_test_split
X=df.iloc[:, :-1]
y=df['LUNG_CANCER']
X_train,X_test,y_train,y_test=train_test_split(X,y,random_state=3,test_size=0.25)

from sklearn.preprocessing import MinMaxScaler
scale=MinMaxScaler()
X_train_scaled=pd.DataFrame(scale.fit_transform(X_train),columns=X_train.columns)
X_train_scaled

```

\	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE
0	0.0	0.575758	0.0	0.0	0.0	1.0
1	0.0	0.696970	1.0	1.0	1.0	1.0
2	0.0	0.712121	0.0	0.0	1.0	1.0
3	0.0	0.575758	0.0	1.0	1.0	0.0
4	1.0	0.651515	1.0	1.0	0.0	1.0
..
226	1.0	1.000000	0.0	0.0	0.0	0.0
227	0.0	0.590909	1.0	0.0	0.0	0.0
228	1.0	0.606061	0.0	1.0	1.0	1.0
229	0.0	0.712121	1.0	0.0	1.0	0.0
230	1.0	0.803030	0.0	1.0	1.0	1.0
\	CHRONIC DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL	CONSUMING
0	0.0	1.0	0.0	1.0		0.0
1	1.0	0.0	0.0	0.0		1.0
2	1.0	0.0	0.0	0.0		1.0
3	0.0	1.0	0.0	1.0		0.0
4	1.0	1.0	1.0	1.0		0.0
..
226	1.0	1.0	0.0	0.0		0.0
227	0.0	1.0	1.0	1.0		1.0
228	0.0	0.0	1.0	1.0		0.0
229	0.0	1.0	0.0	0.0		0.0
230	1.0	1.0	0.0	1.0		1.0
	COUGHING	SHORTNESS OF BREATH	SWALLOWING	DIFFICULTY	CHEST PAIN	

0	0.0	1.0	0.0	1.0
1	0.0	0.0	1.0	1.0
2	0.0	0.0	0.0	0.0
3	0.0	0.0	1.0	1.0
4	1.0	1.0	0.0	0.0
..
226	0.0	1.0	0.0	0.0
227	1.0	1.0	0.0	1.0
228	1.0	0.0	1.0	0.0
229	0.0	0.0	0.0	0.0
230	0.0	0.0	0.0	0.0

[231 rows x 15 columns]

X_test_scaled=pd.DataFrame(scale.fit_transform(X_test),columns=X_test.columns)

X_test_scaled

	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE
\						
0	1.0	0.702703	0.0	0.0	0.0	1.0
1	1.0	0.081081	1.0	1.0	0.0	1.0
2	1.0	0.297297	1.0	0.0	0.0	1.0
3	0.0	0.621622	0.0	0.0	0.0	1.0
4	0.0	0.540541	1.0	0.0	0.0	0.0
..
73	0.0	0.324324	0.0	0.0	0.0	0.0
74	0.0	0.648649	1.0	0.0	0.0	1.0
75	0.0	0.243243	0.0	0.0	0.0	0.0
76	1.0	0.324324	0.0	0.0	0.0	0.0

77	0.0	0.513514	1.0	1.0	1.0	0.0
	CHRONIC DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL	
0	1.0	0.0	1.0	0.0	1.0	
1	1.0	1.0	1.0	1.0	0.0	
2	1.0	1.0	1.0	1.0	1.0	
3	0.0	1.0	0.0	1.0	0.0	
4	0.0	0.0	1.0	1.0	1.0	
..	
73	1.0	1.0	1.0	0.0	1.0	
74	1.0	1.0	1.0	1.0	1.0	
75	1.0	1.0	1.0	0.0	1.0	
76	1.0	0.0	0.0	1.0	0.0	
77	0.0	0.0	0.0	0.0	1.0	

	COUGHING	SHORTNESS OF BREATH	SWALLOWING DIFFICULTY	CHEST PAIN
0	1.0	1.0	0.0	0.0
1	1.0	1.0	0.0	0.0
2	0.0	0.0	1.0	1.0
3	1.0	1.0	0.0	1.0
4	1.0	0.0	0.0	1.0
..
73	1.0	1.0	0.0	1.0
74	1.0	1.0	1.0	1.0
75	0.0	1.0	0.0	1.0
76	0.0	1.0	1.0	0.0
77	0.0	0.0	1.0	1.0

[78 rows x 15 columns]

```
from sklearn.ensemble import RandomForestClassifier
model=RandomForestClassifier()
model.fit(X_train,y_train)
RandomForestClassifier()
prediction=model.predict(X_test)
```

```
prediction
array([1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1,
      1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0,
1,
      1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0,
1,
      1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1], dtype=int64)
from sklearn import metrics
accuracy=metrics.accuracy_score(y_test,prediction)
accuracy
0.9102564102564102
```