In [1]:
```
# https://www.analyticsvidhya.com/blog/2020/10/all-about-decision-tree-from-scratch-with-python-implementation/
```

Build Decision Tree Classifier

In [2]:
```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics
from sklearn.tree import export_text
from sklearn import tree
import matplotlib.pyplot as plt
```

In [3]:
```
columns = ['Blood Pressure', 'diabetes', 'smoking', 'hypertension', 'heart disease']
data = [
    [0, 1, 0, 1, 0],
    [1, 0, 1, 0, 1],
    [1, 1, 0, 0, 1],
    [0, 0, 1, 1, 0],
    [1, 1, 1, 0, 1],
    [0, 1, 1, 0, 0],
    [1, 0, 0, 1, 0],
    [0, 0, 0, 1, 0],
    [1, 1, 0, 1, 1],
    [0, 1, 1, 1, 0],
    [1, 1, 1, 1, 1],
    [0, 0, 1, 0, 0],
    [1, 0, 0, 0, 1],
    [1, 1, 0, 0, 1],
    [0, 1, 0, 0, 0],
    [0, 0, 0, 0, 0],
    [1, 1, 1, 0, 1],
    [0, 1, 1, 0, 0],
    [1, 0, 0, 1, 0],
    [0, 0, 1, 1, 0]
]
```

In [4]:
```
df = pd.DataFrame(data, columns=columns)
df.head()
```

Out[4]:

	Blood Pressure	diabetes	smoking	hypertension	heart disease
0	0	1	0	1	0
1	1	0	1	0	1
2	1	1	0	0	1
3	0	0	1	1	0
4	1	1	1	0	1

In [5]:
```
df_symptoms_X = df.iloc[:,:-1]
df_symptoms_X.head()
```

Out[5]:

	Blood Pressure	diabetes	smoking	hypertension
0	0	1	0	1
1	1	0	1	0
2	1	1	0	0
3	0	0	1	1
4	1	1	1	0

In [6]:
```
df_target_Y = df[['heart disease']]
df_target_Y.head()
```

Out[6]:

	heart disease
0	0
1	1
2	1
3	0
4	1

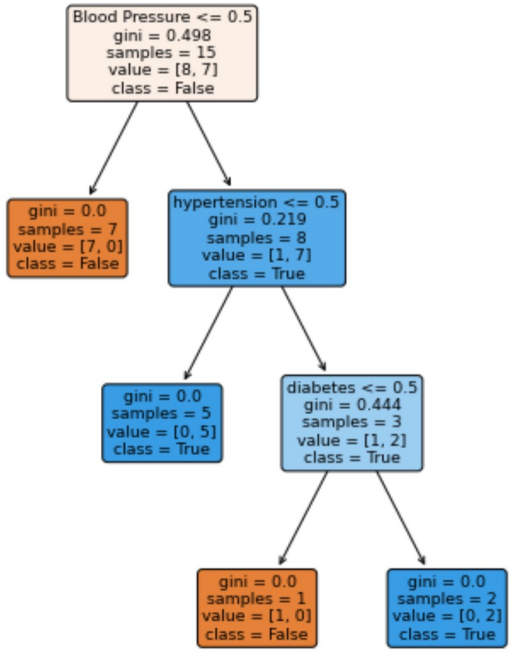
In [7]:
```
CLF = DecisionTreeClassifier()
x_train, x_test, y_train, y_test = train_test_split(df_symptoms_X, df_target_Y)

CLF = CLF.fit(x_train, y_train)
Y_predicted = CLF.predict(x_test)
print("Accuracy = ", metrics.accuracy_score(y_test, Y_predicted))
```

```
Accuracy =  1.0
```

In [8]:
```
print(export_text(CLF, feature_names=columns[:-1]))
```

```
|--- Blood Pressure <= 0.50
|   |--- class: 0
|--- Blood Pressure >  0.50
|   |--- hypertension <= 0.50
|   |   |--- class: 1
|   |--- hypertension >  0.50
|   |   |--- diabetes <= 0.50
|   |   |   |--- class: 0
|   |   |--- diabetes >  0.50
|   |   |   |--- class: 1
```

In [9]:
```
fig = plt.figure(figsize=(6,8))

feature_names = df.columns[:4]
target_names = list(map(lambda x : f"False" if x==0 else f"True", df['heart disease'].unique().tolist()))

_ = tree.plot_tree(CLF,
                   feature_names = feature_names,
                   class_names = target_names,
                   filled=True,
                   rounded = True)
```



In [ ]:

Decision Tree with Lung cancer dataset..

In [10]:
```
df = pd.read_csv("lung_cancer.csv")
print(df.shape)
df.head()
```

```
(309, 16)
```

Out[10]:

	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	CHRONIC DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL CONSUMING	COUGHING	SHORTNESS OF BREATH	SWALLOWING DIFFICULTY	CHI P/
0	M	69	1	2	2	1	1	2	1	2	2	2	2	2	
1	M	74	2	1	1	1	2	2	2	1	1	1	2	2	
2	F	59	1	1	1	2	1	2	1	2	1	2	2	1	
3	M	63	2	2	2	1	1	1	1	1	2	1	1	2	
4	F	63	1	2	1	1	1	1	1	2	1	2	2	1	

Pre-processing data

2 = Yes, 1 = No

In [11]:
```
# df['year'] = np.where(df["year"] > 2016, 1, 0)
for each in df.columns[2:-1]:
    df[each] = np.where(df[each] == 2, 1, 0)

df.head()
```

Out[11]:

	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	CHRONIC DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL CONSUMING	COUGHING	SHORTNESS OF BREATH	SWALLOWING DIFFICULTY	CHI P/
0	M	69	0	1	1	0	0	1	0	1	1	1	1	1	
1	M	74	1	0	0	0	1	1	1	0	0	0	1	1	
2	F	59	0	0	0	1	0	1	0	1	0	1	1	0	
3	M	63	1	1	1	0	0	0	0	0	1	0	0	1	
4	F	63	0	1	0	0	0	0	0	1	0	1	1	0	

In [12]:
```
df_symptoms_X = df.iloc[:,1:-1]
df_symptoms_X.head()
```

Out[12]:

	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	CHRONIC DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL CONSUMING	COUGHING	SHORTNESS OF BREATH	SWALLOWING DIFFICULTY	CHEST PAIN
0	69	0	1	1	0	0	1	0	1	1	1	1	1	1
1	74	1	0	0	0	1	1	1	0	0	0	1	1	1
2	59	0	0	0	1	0	1	0	1	0	1	1	0	1
3	63	1	1	1	0	0	0	0	0	1	0	0	1	1
4	63	0	1	0	0	0	0	0	1	0	1	1	0	0

In [13]:
```
df_target_Y = df[['LUNG_CANCER']]
df_target_Y.head()
```

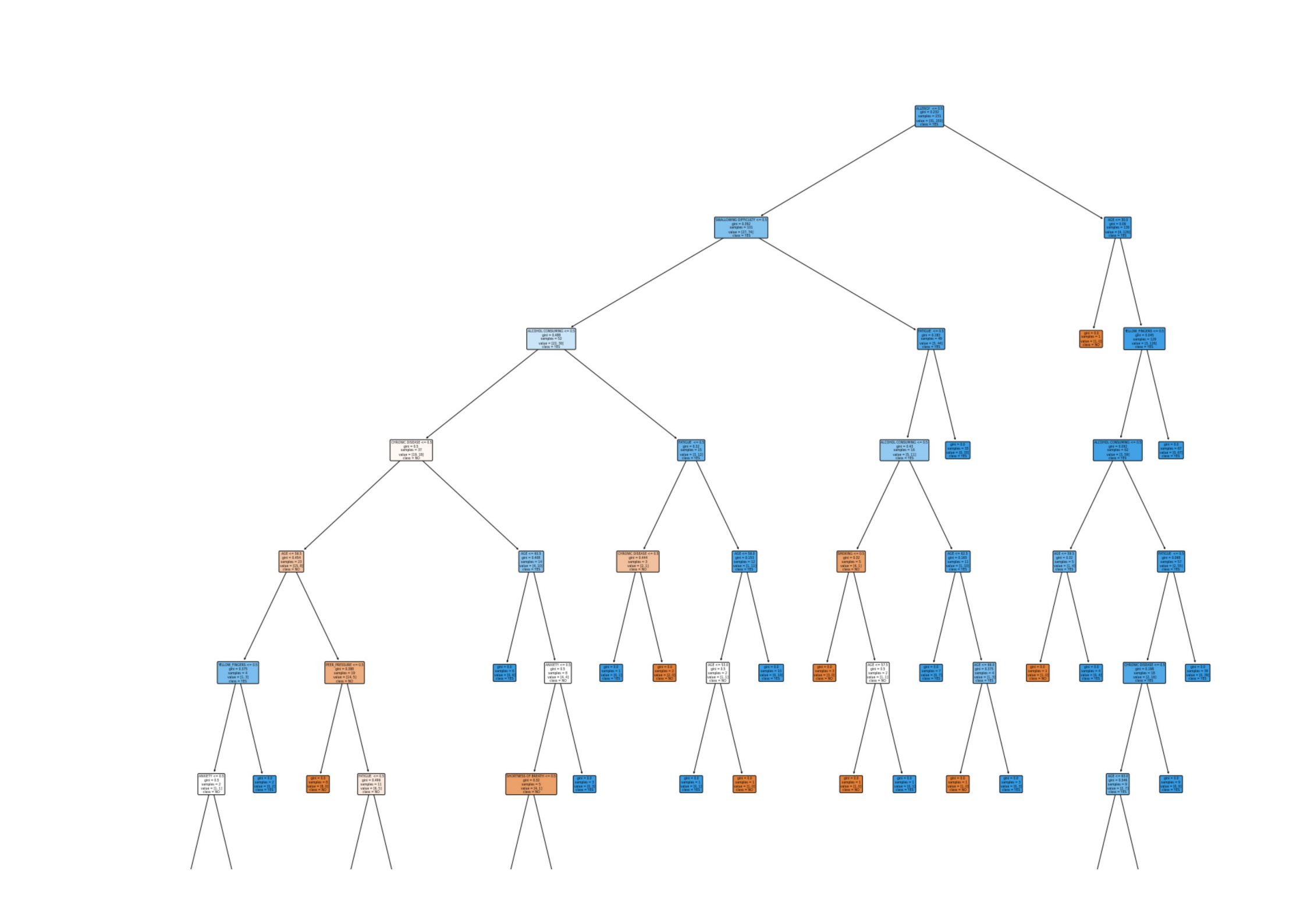
Out[13]:

	LUNG_CANCER
0	YES
1	YES
2	NO
3	NO
4	NO

In [14]:
```
CLF = DecisionTreeClassifier()
x_train, x_test, y_train, y_test = train_test_split(df_symptoms_X, df_target_Y)

CLF = CLF.fit(x_train, y_train)
Y_predicted = CLF.predict(x_test)
print("Accuracy = ", metrics.accuracy_score(y_test, Y_predicted))
```

```
Accuracy =  0.9487179487179487
```

In [22]:
```
fig = plt.figure(figsize=(28,30))

feature_names = df.columns[1:-1]
target_names = sorted(df['LUNG_CANCER'].unique().tolist())

_ = tree.plot_tree(CLF,
                   feature_names = feature_names,
                   class_names = target_names,
                   filled=True,
                   rounded = True)
```


ALLERGY <= 0.5
gini = 0.242
samples = 201
value = [31, 230]
class = YES
WHEEZING DIFFICULTY <= 0.5
gini = 0.382
samples = 120
value = [27, 74]
class = YES
AGE <= 57.5
gini = 0.04
samples = 136
value = [4, 128]
class = YES
ALCOHOL CONSUMING <= 0.5
gini = 0.488
samples = 52
value = [22, 30]
class = YES
FATIGUE <= 0.5
gini = 0.183
samples = 68
value = [5, 60]
class = YES
gini = 0.0
samples = 1
value = [1, 0]
class = NO
YELLOW FINGERS <= 0.5
gini = 0.045
samples = 135
value = [3, 128]
class = YES
CHRONIC DISEASE <= 0.5
gini = 0.5
samples = 37
value = [19, 18]
class = NO
FATIGUE <= 0.5
gini = 0.32
samples = 15
value = [3, 12]
class = YES
ALCOHOL CONSUMING <= 0.5
gini = 0.43
samples = 16
value = [5, 11]
class = YES
gini = 0.0
samples = 52
value = [0, 52]
class = YES
ALCOHOL CONSUMING <= 0.5
gini = 0.062
samples = 62
value = [3, 59]
class = YES
gini = 0.0
samples = 67
value = [0, 67]
class = YES
AGE <= 38.0
gini = 0.454
samples = 23
value = [15, 8]
class = NO
AGE <= 62.0
gini = 0.408
samples = 14
value = [4, 10]
class = YES
CHRONIC DISEASE <= 0.5
gini = 0.444
samples = 3
value = [2, 1]
class = NO
FALL <= 0.5
gini = 0.153
samples = 12
value = [1, 11]
class = YES
SMOKING <= 0.5
gini = 0.32
samples = 5
value = [4, 1]
class = NO
FALL <= 62.0
gini = 0.165
samples = 11
value = [1, 10]
class = YES
AGE <= 59.5
gini = 0.32
samples = 5
value = [1, 4]
class = YES
FATIGUE <= 0.5
gini = 0.069
samples = 57
value = [2, 55]
class = YES
YELLOW FINGERS <= 0.5
gini = 0.375
samples = 4
value = [1, 3]
class = YES
PEER PRESSURE <= 0.5
gini = 0.385
samples = 10
value = [14, 5]
class = NO
gini = 0.0
samples = 6
value = [6, 0]
class = NO
ANXIETY <= 0.5
gini = 0.5
samples = 8
value = [4, 4]
class = NO
gini = 0.0
samples = 2
value = [2, 0]
class = NO
gini = 0.0
samples = 1
value = [0, 1]
class = YES
AGE <= 57.0
gini = 0.5
samples = 2
value = [1, 1]
class = NO
gini = 0.0
samples = 10
value = [0, 10]
class = YES
gini = 0.0
samples = 1
value = [1, 0]
class = NO
AGE <= 57.5
gini = 0.5
samples = 2
value = [1, 1]
class = NO
gini = 0.0
samples = 3
value = [0, 3]
class = YES
AGE <= 66.0
gini = 0.375
samples = 4
value = [1, 3]
class = YES
gini = 0.0
samples = 1
value = [1, 0]
class = NO
gini = 0.0
samples = 4
value = [0, 4]
class = YES
gini = 0.0
samples = 1
value = [1, 0]
class = NO
CHRONIC DISEASE <= 0.5
gini = 0.198
samples = 10
value = [1, 16]
class = YES
gini = 0.0
samples = 47
value = [0, 47]
class = YES
ANXIETY <= 0.5
gini = 0.5
samples = 2
value = [1, 1]
class = NO
gini = 0.0
samples = 2
value = [0, 2]
class = YES
FATIGUE <= 0.5
gini = 0.496
samples = 11
value = [6, 5]
class = NO
SHORTNESS OF BREATH <= 0.5
gini = 0.32
samples = 5
value = [4, 1]
class = NO
gini = 0.0
samples = 1
value = [0, 3]
class = YES
gini = 0.0
samples = 1
value = [1, 0]
class = NO
gini = 0.0
samples = 1
value = [0, 1]
class = YES
gini = 0.0
samples = 1
value = [1, 0]
class = NO
gini = 0.0
samples = 1
value = [0, 1]
class = YES
gini = 0.0
samples = 1
value = [1, 0]
class = NO
gini = 0.0
samples = 1
value = [0, 1]
class = YES
gini = 0.0
samples = 1
value = [1, 0]
class = NO
gini = 0.0
samples = 1
value = [0, 1]
class = YES
AGE <= 63.0
gini = 0.346
samples = 9
value = [2, 7]
class = YES
gini = 0.0
samples = 1
value = [0, 9]
class = YES

In [ ]:

In [16]:
```
print(export_text(CLF,feature_names=list(feature_names)))
```

```
|--- ALLERGY  <= 0.50
|   |--- SWALLOWING DIFFICULTY <= 0.50
|   |   |--- ALCOHOL CONSUMING <= 0.50
|   |   |   |--- CHRONIC DISEASE <= 0.50
|   |   |   |   |--- AGE <= 56.50
|   |   |   |   |   |--- YELLOW_FINGERS <= 0.50
|   |   |   |   |   |   |--- ANXIETY <= 0.50
|   |   |   |   |   |   |   |--- class: YES
|   |   |   |   |   |   |--- ANXIETY >  0.50
|   |   |   |   |   |   |   |--- class: NO
|   |   |   |   |   |--- YELLOW_FINGERS >  0.50
|   |   |   |   |   |   |--- class: YES
|   |   |   |   |--- AGE >  56.50
|   |   |   |   |   |--- PEER_PRESSURE <= 0.50
|   |   |   |   |   |   |--- class: NO
|   |   |   |   |   |--- PEER_PRESSURE >  0.50
|   |   |   |   |   |   |--- FATIGUE  <= 0.50
|   |   |   |   |   |   |   |--- class: NO
|   |   |   |   |   |   |--- FATIGUE  >  0.50
|   |   |   |   |   |   |   |--- AGE <= 60.00
|   |   |   |   |   |   |   |   |--- class: NO
|   |   |   |   |   |   |   |--- AGE >  60.00
|   |   |   |   |   |   |   |   |--- AGE <= 68.50
|   |   |   |   |   |   |   |   |   |--- class: YES
|   |   |   |   |   |   |   |   |--- AGE >  68.50
|   |   |   |   |   |   |   |   |   |--- class: NO
|   |   |   |--- CHRONIC DISEASE >  0.50
|   |   |   |   |--- AGE <= 60.50
|   |   |   |   |   |--- class: YES
|   |   |   |   |--- AGE >  60.50
|   |   |   |   |   |--- ANXIETY <= 0.50
|   |   |   |   |   |   |--- SHORTNESS OF BREATH <= 0.50
|   |   |   |   |   |   |   |--- SMOKING <= 0.50
|   |   |   |   |   |   |   |   |--- class: YES
|   |   |   |   |   |   |   |--- SMOKING >  0.50
|   |   |   |   |   |   |   |   |--- class: NO
|   |   |   |   |   |   |--- SHORTNESS OF BREATH >  0.50
|   |   |   |   |   |   |   |--- class: NO
|   |   |   |   |   |--- ANXIETY >  0.50
|   |   |   |   |   |   |--- class: YES
|   |   |--- ALCOHOL CONSUMING >  0.50
|   |   |   |--- FATIGUE  <= 0.50
|   |   |   |   |--- CHRONIC DISEASE <= 0.50
|   |   |   |   |   |--- class: YES
|   |   |   |   |--- CHRONIC DISEASE >  0.50
|   |   |   |   |   |--- class: NO
|   |   |   |--- FATIGUE  >  0.50
|   |   |   |   |--- AGE <= 56.00
|   |   |   |   |   |--- AGE <= 53.00
|   |   |   |   |   |   |--- class: YES
```

```
|   |   |   |   |   |--- AGE >  53.00
|   |   |   |   |   |   |--- class: NO
|   |   |   |   |   |--- AGE >  56.00
|   |   |   |   |   |--- class: YES
|   |--- SWALLOWING DIFFICULTY >  0.50
|   |   |--- FATIGUE  <= 0.50
|   |   |   |--- ALCOHOL CONSUMING <= 0.50
|   |   |   |   |--- SMOKING <= 0.50
|   |   |   |   |   |--- class: NO
|   |   |   |   |--- SMOKING >  0.50
|   |   |   |   |   |--- AGE <= 57.50
|   |   |   |   |   |   |--- class: NO
|   |   |   |   |   |--- AGE >  57.50
|   |   |   |   |   |   |--- class: YES
|   |   |   |--- ALCOHOL CONSUMING >  0.50
|   |   |   |   |--- AGE <= 62.50
|   |   |   |   |   |--- class: YES
|   |   |   |   |--- AGE >  62.50
|   |   |   |   |   |--- AGE <= 66.00
|   |   |   |   |   |   |--- class: NO
|   |   |   |   |   |--- AGE >  66.00
|   |   |   |   |   |   |--- class: YES
|   |   |--- FATIGUE  >  0.50
|   |   |   |--- class: YES
|--- ALLERGY  >  0.50
|   |--- AGE <= 30.00
|   |   |--- class: NO
|   |--- AGE >  30.00
|   |   |--- YELLOW_FINGERS <= 0.50
|   |   |   |--- ALCOHOL CONSUMING <= 0.50
|   |   |   |   |--- AGE <= 59.50
|   |   |   |   |   |--- class: NO
|   |   |   |   |--- AGE >  59.50
|   |   |   |   |   |--- class: YES
|   |   |   |--- ALCOHOL CONSUMING >  0.50
|   |   |   |   |--- FATIGUE  <= 0.50
|   |   |   |   |   |--- CHRONIC DISEASE <= 0.50
|   |   |   |   |   |   |--- AGE <= 63.00
|   |   |   |   |   |   |   |--- class: YES
|   |   |   |   |   |   |--- AGE >  63.00
|   |   |   |   |   |   |   |--- AGE <= 69.50
|   |   |   |   |   |   |   |   |--- AGE <= 66.50
|   |   |   |   |   |   |   |   |   |--- class: NO
|   |   |   |   |   |   |   |   |--- AGE >  66.50
|   |   |   |   |   |   |   |   |   |--- class: NO
|   |   |   |   |   |   |   |--- AGE >  69.50
|   |   |   |   |   |   |   |   |--- class: YES
|   |   |   |   |   |--- CHRONIC DISEASE >  0.50
|   |   |   |   |   |   |--- class: YES
|   |   |   |   |--- FATIGUE  >  0.50
```

```
|   |   |   |   |   |--- class: YES
|   |   |--- YELLOW_FINGERS >  0.50
|   |   |   |--- class: YES
```

In [ ]: