```
In [50]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sn
```

In [51]: path = 'https://raw.githubusercontent.com/ovibaridar/Data_sets/main/cancer%20pa

In [52]: data = pd.read_csv(path)

In [53]: data.head()

Out[53]:

	index	Patient Id	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	OccuPational Hazards	Genetic Risk	chronic Lung Disease	
0	0	P1	33	1	2	4	5	4	3	2	
1	1	P10	17	1	3	1	5	3	4	2	
2	2	P100	35	1	4	5	6	5	5	4	
3	3	P1000	37	1	7	7	7	7	6	7	
4	4	P101	46	1	6	8	7	7	7	6	

5 rows × 26 columns

In [54]: data = data.drop(['index','Patient Id'], axis=1)

In [55]: data.head()

Out[55]:

	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	OccuPational Hazards	Genetic Risk	chronic Lung Disease	Balanced Diet	Obesity
0	33	1	2	4	5	4	3	2	2	4
1	17	1	3	1	5	3	4	2	2	2
2	35	1	4	5	6	5	5	4	6	7
3	37	1	7	7	7	7	6	7	7	7
4	46	1	6	8	7	7	7	6	7	7

5 rows × 24 columns

```
In [56]: data.info()
```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 24 columns):

#	Column	Non-Null Count	Dtype
0	Age	1000 non-null	int64
1	Gender	1000 non-null	int64
2	Air Pollution	1000 non-null	int64
3	Alcohol use	1000 non-null	int64
4	Dust Allergy	1000 non-null	int64
5	OccuPational Hazards	1000 non-null	int64
6	Genetic Risk	1000 non-null	int64
7	chronic Lung Disease	1000 non-null	int64
8	Balanced Diet	1000 non-null	int64
9	Obesity	1000 non-null	int64
10	Smoking	1000 non-null	int64
11	Passive Smoker	1000 non-null	int64
12	Chest Pain	1000 non-null	int64
13	Coughing of Blood	1000 non-null	int64
14	Fatigue	1000 non-null	int64
15	Weight Loss	1000 non-null	int64
16	Shortness of Breath	1000 non-null	int64
17	Wheezing	1000 non-null	int64
18	Swallowing Difficulty	1000 non-null	int64
19	Clubbing of Finger Nails	1000 non-null	int64
20	Frequent Cold	1000 non-null	int64
21	Dry Cough	1000 non-null	int64
22	Snoring	1000 non-null	int64
23	Level	1000 non-null	object
	1		

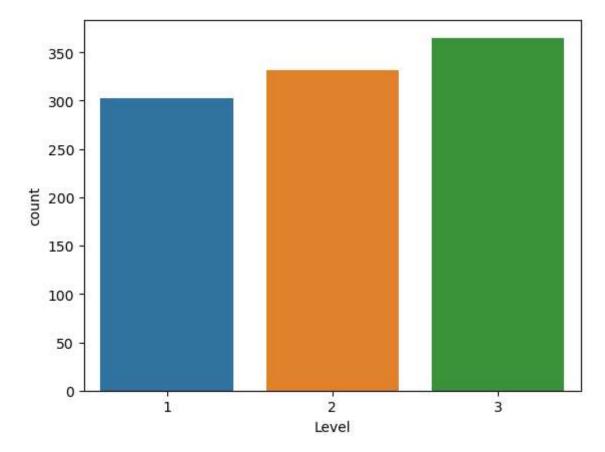
dtypes: int64(23), object(1)
memory usage: 187.6+ KB

```
In [57]: data.isnull().sum()
Out[57]: Age
                                      0
                                       0
         Gender
         Air Pollution
                                       0
         Alcohol use
                                       0
         Dust Allergy
                                       0
         OccuPational Hazards
                                       0
                                       0
         Genetic Risk
         chronic Lung Disease
                                      0
         Balanced Diet
                                       0
         Obesity
                                       0
         Smoking
                                       0
                                       0
         Passive Smoker
         Chest Pain
                                      0
         Coughing of Blood
                                       0
         Fatigue
                                       0
         Weight Loss
                                       0
         Shortness of Breath
                                       0
         Wheezing
                                      0
                                       0
         Swallowing Difficulty
         Clubbing of Finger Nails
                                      0
         Frequent Cold
                                       0
                                       0
         Dry Cough
         Snoring
                                       0
         Level
                                       0
         dtype: int64
In [58]:
         data.shape
Out[58]: (1000, 24)
In [59]: data.columns
Out[59]: Index(['Age', 'Gender', 'Air Pollution', 'Alcohol use', 'Dust Allergy',
                 'OccuPational Hazards', 'Genetic Risk', 'chronic Lung Disease',
                 'Balanced Diet', 'Obesity', 'Smoking', 'Passive Smoker', 'Chest Pain',
                 'Coughing of Blood', 'Fatigue', 'Weight Loss', 'Shortness of Breath',
                 'Wheezing', 'Swallowing Difficulty', 'Clubbing of Finger Nails',
                 'Frequent Cold', 'Dry Cough', 'Snoring', 'Level'],
                dtype='object')
In [60]: | data['Level'] = data['Level'].replace({'Low': 1, 'Medium': 2, 'High': 3})
In [61]: data['Level'].value_counts()
Out[61]: Level
         3
               365
         2
               332
               303
         Name: count, dtype: int64
```

grafical view

```
In [62]: sn.countplot(data= data ,x='Level')
```

Out[62]: <Axes: xlabel='Level', ylabel='count'>



```
cancer patient data sets - Jupyter Notebook
                plt.figure(figsize=(10,10))
In [63]:
                sn.heatmap(data.corr() ,annot = True)
Out[63]: <Axes: >
                                                                                                                                                     - 1.0
                                               -0.20.099.15.035062207 B. 1220049934005004991280532095.10.0355.095.10.0398.0132004006
                                                   0.250.230.20.190.230.210.10.130.250.180.230.150.142.0538043607360538403400653140.140.1
                              Alcohol use -0.150.2-0.75 1 0.820.880.880.760.650.670.550.590.720.670.240.210.440.180.110.410.180.210.120.72
                                                                                                                                                    - 0.8
                              Dust Allergy -,0350.20.6-0.82 1 0.840.750.620.650.70.360.560.640.630.330.320.520.30.0310.350.220.30.056
                    OccuPational Hazards -,0620.19.6 0.880.84 1 0.890.860.650.72 0.50.560.760.650.270.180.370.180.002930.070.16.026.6
                              Genetic Risk -.07-3.220.7 0.880.790.89 1 0.840.6 0.7 0.540.6 10.8 0.6 0.230.2 70.460.20.060 36.08 0.140.05 7
                    chronic Lung Disease -0.130.2 D.650.760.620.860.84 1 0.620.60.580.570.780.60.250.10.18.06.0070.0078.30.029.1 D.048.6
                                                                                                                                                     - 0.6
                            Balanced Diet -0049.10.520.650.650.650.690.680.62 1 0.710.650.730.80.75 0.40.000534.00040470470470.260.330.150
                                   Obesity -.034.120.60.670.70.720.730.60.71 1 0.490.680.670.810.550.310.410.094.130.150.290.20.030.83
                                            .07-<u>1</u>0.210.480.550.360.50.540.580.650.49<mark>.1.0.76</mark>0.650.560.2-0.201.02080407.240.040.040.010.190.5
                                  Smoking
                                            -0049.180.610.590.560.560.610.570.730.630.76 1 0.70.640.38.05380630.20.345.0346.10.120.25
                          Passive Smoker
                                                                                                                                                     0.4
                                            -.01-3.270.590.720.640.780.830.780.80.670.650.7 1 0.710.205.0001240.10.0702080.040.140.140
                                            -.05-19.150.610.670.630.650.630.60.750.810.560.640.71 1 0.480.110.340.0860805.0605.240.110.080.78
                       Coughing of Blood
                                   Fatigue -.090.120.210.240.330.270.230.250.40.550.20.380.250.48 1 0.470.40.170.16.0410.410.270.230
                              Weight Loss -0.1-0.05@260.210.320.180.270-0.00@5310.20.0580010110.47 1 0.570.36.058.380.160.190.19.35
                                                                                                                                                     0.2
                      Shortness of Breath -.035.045.270.440.520.370.460.180.340.400.020068.240.320.40.57 1 0.21-0.20.470.350.490.160.5
                                 Wheezing -).09050706056.180.30.180.20.050706040904.0470.20.1-0.0805170.330.21 1.0.390.39.0909050.120.2-
                     Swallowing Difficulty -0.10.058080.10.0810009068003040.130.240.36.072086.16.0530.20.39 1 0.120.18.055.210.25
                                                                                                                                                     - 0.0
                  Clubbing of Finger Nails -.039.034.240.410.350.370.360.30.040.1-10.0411.036080.0660410.380.470.340.12 1 0.240.3-10.0168.24
                            Frequent Cold -).00100005.170.180.20.0707080029.260.290.040.10.048.240.410.160.35.099.130.24 1 0.520.340.4
                               Dry Cough -.01-2.12.26.210.30.16.19.110.330.20.010.120.140.150.270.190.49.054.055.310.52 1 0.180.3
                                   Snoring -0047/18.020.12.05780278.057048.16.039.190.250.19.088.230.190.16.120.2-0.018.340.18 1
                                                                               Obesity
                                                                                                              Wheezing
                                                                                   Smoking
                                                                                                   Fatigue
                                                    Air Pollution
                                                        Alcohol use
                                                                OccuPational Hazards
                                                                       chronic Lung Disease
                                                                           Balanced Diet
                                                                                           Chest Pain
                                                                                              Coughing of Blood
                                                                                                      Weight Loss
                                                                                                                  Swallowing Difficulty
                                                                                                                      Clubbing of Finger Nails
                                                                                                                          Frequent Cold
                                                            Dust Allergy
                                                                    Genetic Risk
                                                                                       Passive Smoker
                                                                                                          Shortness of Breath
                                                                                                                              Dry Cough
                from sklearn.model selection import train test split
                x= data.drop('Level', axis=1)
In [65]:
```

```
In [64]: from sklearn.model_selection import train_test_split

In [65]: x= data.drop('Level', axis=1)
y= data.Level

In [66]: xtrain,xtest,ytrain,ytest = train_test_split(x,y,test_size = .30,random_state = .30]
In [67]: xtrain.shape

Out[67]: (700, 23)
```

```
xtest.shape
In [68]:
Out[68]: (300, 23)
In [69]:
                            plt.figure(figsize=(10,10))
                            sn.heatmap(xtrain.corr() ,annot = True)
Out[69]: <Axes: >
                                                                                                                                                                                                                                                             - 1.0
                                                                  Age - 1 0.18.09 B.1 D.01030 5030 5 D.1 B.02040 4 D.1 D.0-D500 0.D6 D.1 D.0 690 0 5 05.1 20.1 Q.0 1 Q.0
                                                                                        -0.2-0.190.160.150.190.140.0706080.190.140.180.1-0.0701.0407.0406040807050207.01-0.120.1
                                                    Air Pollution - 0930.2
                                                                                          1 0.750.63 0.6 0.7 0.630.53 0.6 0.480.610.580.610.220.280.28.036.079.230.170.24.04
                                                     Alcohol use -0.150.10.75 1 0.820.880.870.770.660.680.560.60.710.680.260.230.430.150.0930.40.180.170.11
                                                                                                                                                                                                                                                             - 0.8
                                                    Dust Allergy -.0180.16.630.82 1 0.840.790.630.660.7 0.360.570.640.630.340.330.530.28.0350.350.220.29.03
                                   OccuPational Hazards -,058.150.60.880.84 1 0.890.86 0.70.73 0.5 0.570.770.650.280.2 0.370.15.019.36.085.14.01
                                                   Genetic Risk -.055.190.70.870.790.89 1 0.850.680.720.540.610.830.630.240.290.460.180.050.350.0850.140.07
                                   chronic Lung Disease -0.130.170.630.770.630.860.85 1 0.640.610.590.590.78 0.6 0.260.130.19.049402 50.3 0.044.110.03
                                                                                                                                                                                                                                                             0.6
                                                Balanced Diet -J.024.076.530.660.660.70.680.64 1 0.710.650.730.80.740.40.028.35.040.050.049.260.310.13
                                                                           -.049.0870.60.68.0,70.730.720.610.71<mark>. 1</mark>30.490.680.66<mark>0.81</mark>0.550.330.410.070.130.170.280.110.02
                                                                          -0.110.140.480.560.360.50.540.590.630.49 1 0.760.630.560.2-0.140.016.080.2-40.0350-D90040217
                                             Passive Smoker -.0150.140.610.60.570.570.610.590.730.680.76 1 0.680.640.38.078086.160.340.018087.098.21
                                                                                                                                                                                                                                                              0.4
                                                       Chest Pain - 00 17, 18, 580, 710, 640, 770, 830, 780, 80, 660, 630, 68 1 0, 70, 25, 00 50, 28, 0 70, 070, 070, 083, 048, 120, 11
                                        Coughing of Blood -.06-70.1 D.610.680.630.650.630.60.740.810.560.640.7 1 0.470.110.310.10.099.060.240.10.06
                                                            Fatigue -0.1-D.070.220.260.340.280.240.260.410.55 0.2 0.380.250.47 1 0.470.420.160.16.07 3.420.270.24
                                                                                                                                                                                                                                                             - 0.2
                                                    Weight Loss -.069.040.280.230.330.20.290.130.028.330.19.0090050.110.47 1 0.590.340.0730.40.180.2-0.19
                                      Shortness of Breath -0.056046.280.430.530.370.460.190.350.440.016086.230.310.420.59 1 0.250.140.460.360.5-0.1
                                                        Wheezing -0.19.048036.150.280.150.18.0444048.070.080.16.0740.110.160.340.25 1 0.390.360.110.080.1
                                     Swallowing Difficulty -0.14.075.075.075.075.092.035.019.051.025.058.130.240.34.071.099.16.0730.140.39 1 0.066.150.046.22
                                                                                                                                                                                                                                                              0.0
                               Clubbing of Finger Nails - 0.010.020.23 0.40.350.360.350.310.049.170.035.013080.06010730.40.460.360.06 1 0.240.30.02
                                                Frequent Cold 4.0404010.170.180.270.085085044.260.280.089087048.240.420.180.360.110.150.24 1 0.490.3
                                                      Dry Cough -0.01-5.120.240.170.290.140.180.110.310.107.0092098.120.10.270.2 0.5 0.040.0430.3 0.49 1
                                                            Snoring -0.01-0.1-0.046.1 0.0305.018.076.030.130.026.170.210.10.0620.240.190.170.1 0.2-20.026.340.15
                                                                                                                                          Obesity
                                                                                                                                                                            Fatigue
                                                                                                                                                                                                             Clubbing of Finger Nails
                                                                                                                     Genetic Risk
                                                                                                                            chronic Lung Disease
                                                                                                                                                       Passive Smoker
                                                                                                                                                              Chest Pain
                                                                                                                                                                     Coughing of Blood
                                                                                                                                                                                  Weight Loss
                                                                                                                                                                                          Shortness of Breath
                                                                                                                                                                                                       Swallowing Difficulty
                                                                                                                                                                                                                     Frequent Cold
                                                                                                                                                                                                                            Dry Cough
                                                                                          Air Pollution
                                                                                                 Alcohol use
                                                                                                        Dust Allergy
                                                                                                               OccuPational Hazards
                                                                                                                                   Balanced Diet
                                                                                                                                                 Smoking
                                                                                                                                                                                                 Wheezing
In [70]:
                            from sklearn.tree import DecisionTreeClassifier
                            dtc = DecisionTreeClassifier()
```

```
In [72]: dtc.fit(xtrain,ytrain)
Out[72]:
          ▼ DecisionTreeClassifier
          DecisionTreeClassifier()
         perd = dtc.predict(xtest)
In [73]:
         perd
Out[73]: array([2, 2, 2, 3, 2, 2, 2, 2, 2, 1, 2, 2, 1, 3, 1, 3, 3, 3, 3, 2,
                1, 2, 3, 1, 2, 3, 1, 1, 1, 3, 2, 1, 1, 3, 3, 2, 3, 1, 3, 3, 1, 2,
                2, 1, 1, 2, 3, 1, 2, 3, 1, 3, 3, 1, 3, 3, 1, 2, 1, 3, 3, 1, 2,
                3, 3, 2, 3, 3, 1, 2, 2, 2, 2, 3, 1, 1, 2, 2, 2, 1, 2, 3, 3, 3, 1,
                3, 3, 1, 1, 2, 3, 3, 3, 3, 3, 2, 1, 1, 3, 1, 1, 2, 3, 3, 2,
                2, 3, 3, 1, 2, 3, 2, 3, 2, 2, 3, 3, 1, 1, 2, 3, 3, 3, 3, 3, 1, 2,
                1, 3, 2, 1, 2, 3, 2, 2, 1, 2, 1, 1, 3, 2, 1, 3, 1, 3, 3, 1, 1, 3,
                1, 3, 3, 3, 1, 3, 1, 3, 3, 3, 3, 2, 2, 3, 3, 3, 3, 2, 3, 3, 1, 3,
                2, 1, 2, 3, 1, 3, 1, 2, 3, 2, 3, 2, 1, 2, 1, 1, 2, 3, 3, 2, 3, 3,
                1, 2, 2, 3, 2, 2, 1, 1, 1, 3, 1, 1, 2, 3, 1, 1, 3, 3, 3, 2, 2, 1,
                2, 3, 3, 1, 2, 3, 3, 3, 2, 3, 3, 1, 3, 1, 3, 2, 2, 2, 3, 3, 3,
                2, 3, 2, 3, 3, 2, 2, 2, 2, 3, 2, 1, 2, 1, 2, 3, 1, 3, 2, 3, 3, 1,
                1, 3, 3, 1, 2, 1, 2, 1, 2, 3, 2, 2, 2, 1, 2, 1, 2, 3, 3, 1, 3, 1,
                2, 3, 3, 1, 1, 2, 2, 3, 2, 1, 1, 3, 1, 1], dtype=int64)
In [74]: | dtc.score(xtrain,ytrain)
Out[74]: 1.0
In [75]: dtc.score(xtest,ytest)
Out[75]: 1.0
In [84]: from sklearn.metrics import r2_score
In [85]: | r2_score = r2_score(ytest,perd)
In [86]: r2_score
Out[86]: 1.0
```