In [1]:

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

In [2]:

```
df=pd.read_excel('Health Data.xlsx')
```

In [3]:

```
df.shape
```

Out[3]:

```
(334, 31)
```

In [4]:

```
df.head()
```

Out[4]:

	I am currently employed at least part-time	I identify as having a mental illness	Education	I have my own computer separate from a smart phone	I have been hospitalized before for my mental illness	How many days were you hospitalized for your mental illness	I am legally disabled	I have my regular access to the internet	p
0	0	0	High School or GED	0	0	0.0	0	1	
1	1	1	Some Phd	1	0	0.0	0	1	
2	1	0	Completed Undergraduate	1	0	0.0	0	1	
3	0	0	Some Undergraduate	1	0	NaN	0	1	
4	1	1	Completed Undergraduate	1	1	35.0	1	1	

5 rows × 31 columns

In [5]:

```
df.tail()
```

Out[5]:

	I am currently employed at least part-time	I identify as having a mental illness	Education	I have my own computer separate from a smart phone	I have been hospitalized before for my mental illness	How many days were you hospitalized for your mental illness	I am legally disabled	I have my regular access to the internet
329	0	0	High School or GED	1	0	NaN	1	1
330	1	0	Some Undergraduate	1	0	0.0	0	1
331	1	0	Some Undergraduate	1	0	0.0	0	1
332	0	1	Some Undergraduate	0	1	1.0	1	1
333	1	1	Some Undergraduate	1	0	0.0	1	1

5 rows × 31 columns

In [6]:

```
columns_to_drop=['Region','I have my regular access to the internet','I am currently employ
```

In [7]:

```
df.drop(columns=columns_to_drop,inplace=True)
```

In [8]:

```
df['Annual income (including any social welfare programs) in Rupee']=df['Annual income (inc
```

In [9]:

```
df.drop('Annual income (including any social welfare programs) in USD',axis=1,inplace=True)
```

In [10]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 334 entries, 0 to 333
Data columns (total 21 columns):
 #   Column                                                  Non-Nul
l Count  Dtype
---  ------                                                  -------
-------  -----
 0   I identify as having a mental illness                   334 non
-null    int64
 1   Education                                               334 non
-null    object
 2   I have my own computer separate from a smart phone      334 non
-null    int64
 3   I have been hospitalized before for my mental illness   334 non
-null    int64
 4   How many days were you hospitalized for your mental illness     297 non
-null    float64
 5   I am legally disabled                                   334 non
-null    int64
 6   I live with my parents                                  334 non
-null    int64
 7   I am unemployed                                         334 non
-null    int64
 8   I read outside of work and school                       334 non
-null    int64
 9   How many times were you hospitalized for your mental illness    334 non
-null    int64
 10  Lack of concentration                                   333 non
-null    float64
 11  Anxiety                                                 334 non
-null    int64
 12  Depression                                              334 non
-null    int64
 13  Obsessive thinking                                      333 non
-null    float64
 14  Mood swings                                             333 non
-null    float64
 15  Panic attacks                                           333 non
-null    float64
 16  Compulsive behavior                                     333 non
-null    float64
 17  Tiredness                                               333 non
-null    float64
 18  Age                                                     334 non
-null    object
 19  Gender                                                  334 non
-null    object
 20  Annual income (including any social welfare programs) in Rupee  334 non
-null    int64
dtypes: float64(7), int64(11), object(3)
memory usage: 54.9+ KB
```

In [11]:

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

Out[11]:

```
I identify as having a mental illness                                0
Education                                                            0
I have my own computer separate from a smart phone                   0
I have been hospitalized before for my mental illness                0
How many days were you hospitalized for your mental illness         37
I am legally disabled                                                0
I live with my parents                                               0
I am unemployed                                                      0
I read outside of work and school                                    0
How many times were you hospitalized for your mental illness         0
Lack of concentration                                                1
Anxiety                                                              0
Depression                                                           0
Obsessive thinking                                                   1
Mood swings                                                          1
Panic attacks                                                        1
Compulsive behavior                                                  1
Tiredness                                                            1
Age                                                                  0
Gender                                                               0
Annual income (including any social welfare programs) in Rupee       0
dtype: int64
```

In [12]:

```
for i in df:
    if i=='Education' or i=='Age' or i=='Gender':
        df[i].dropna()
    else:
        df[i].fillna(df[i].median(),inplace=True)
```

In [13]:

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

Out[13]:

```
I identify as having a mental illness                                0
Education                                                            0
I have my own computer separate from a smart phone                   0
I have been hospitalized before for my mental illness                0
How many days were you hospitalized for your mental illness          0
I am legally disabled                                                0
I live with my parents                                               0
I am unemployed                                                      0
I read outside of work and school                                    0
How many times were you hospitalized for your mental illness         0
Lack of concentration                                                0
Anxiety                                                              0
Depression                                                           0
Obsessive thinking                                                   0
Mood swings                                                          0
Panic attacks                                                        0
Compulsive behavior                                                  0
Tiredness                                                            0
Age                                                                  0
Gender                                                               0
Annual income (including any social welfare programs) in Rupee       0
dtype: int64
```

In [14]:

```
y=df['I identify as having a mental illness']
```

In [15]:

```
df.drop('I identify as having a mental illness',axis=1,inplace=True)
```

In [16]:

```
df.head(20)
```

Out[16]:

	Education	I have my own computer separate from a smart phone	I have been hospitalized before for my mental illness	How many days were you hospitalized for your mental illness	I am legally disabled	I live with my parents	I am unemployed	I read outside of work and school
0	High School or GED	0	0	0.0	0	0	1	1
1	Some Phd	1	0	0.0	0	0	0	1
2	Completed Undergraduate	1	0	0.0	0	0	0	1
3	Some Undergraduate	1	0	0.0	0	1	1	1
4	Completed Undergraduate	1	1	35.0	1	0	0	1
5	High School or GED	1	0	0.0	0	1	0	1
6	Some Undergraduate	1	0	0.0	0	0	0	1
7	Some Undergraduate	1	0	0.0	0	1	0	1
8	Completed Undergraduate	1	0	0.0	0	0	0	1
9	Some Masters	1	0	0.0	0	0	0	1
10	Completed Undergraduate	1	0	0.0	0	0	1	1
11	Completed Undergraduate	1	0	0.0	0	0	0	1
12	Completed Undergraduate	1	0	0.0	0	0	0	1
13	Completed Masters	1	0	0.0	0	0	0	1
14	High School or GED	1	0	0.0	0	0	0	1
15	Completed Undergraduate	1	0	0.0	0	0	0	1
16	Some Masters	1	1	65.0	1	1	1	0
17	Some Undergraduate	1	0	0.0	0	0	0	0
18	Some Masters	1	0	0.0	0	0	0	1
19	Completed Phd	0	0	0.0	0	0	1	1

In [17]:

```
df.corr()
```

Out[17]:

	I have my own computer separate from a smart phone	I have been hospitalized before for my mental illness	How many days were you hospitalized for your mental illness	I am legally disabled	I live with my parents	I am unemployed	I ou of sc
I have my own computer separate from a smart phone	1.000000	-0.125730	-0.133132	-0.116516	-0.125075	-0.210350	0.00
I have been hospitalized before for my mental illness	-0.125730	1.000000	0.532231	0.315741	0.075473	0.186704	0.03
How many days were you hospitalized for your mental illness	-0.133132	0.532231	1.000000	0.241376	0.141764	0.134231	-0.04
I am legally disabled	-0.116516	0.315741	0.241376	1.000000	0.138883	0.332804	0.02
I live with my parents	-0.125075	0.075473	0.141764	0.138883	1.000000	0.163039	-0.14
I am unemployed	-0.210350	0.186704	0.134231	0.332804	0.163039	1.000000	-0.03
I read outside of work and school	0.009992	0.031341	-0.043065	0.020962	-0.148967	-0.032137	1.00
How many times were you hospitalized for your mental illness	-0.016522	0.359856	0.583789	0.073773	0.116312	0.062755	-0.02
Lack of concentration	-0.039837	0.156268	-0.028440	0.166285	0.115374	0.092676	0.09
Anxiety	-0.205544	0.249252	0.176289	0.112163	0.165007	0.168219	0.02
Depression	-0.107092	0.340045	0.219452	0.241015	0.163039	0.170011	0.05
Obsessive thinking	-0.128506	0.361651	0.203541	0.177039	0.153846	0.148395	0.07
Mood swings	-0.176926	0.247821	0.127459	0.165751	0.234040	0.134034	-0.02
Panic attacks	-0.097961	0.321701	0.257111	0.146293	0.123265	0.220301	0.03
Compulsive behavior	-0.075467	0.267582	0.167823	0.147335	0.196067	0.037277	0.10
Tiredness	-0.008382	0.151655	0.000335	0.090255	-0.001621	0.063563	0.10

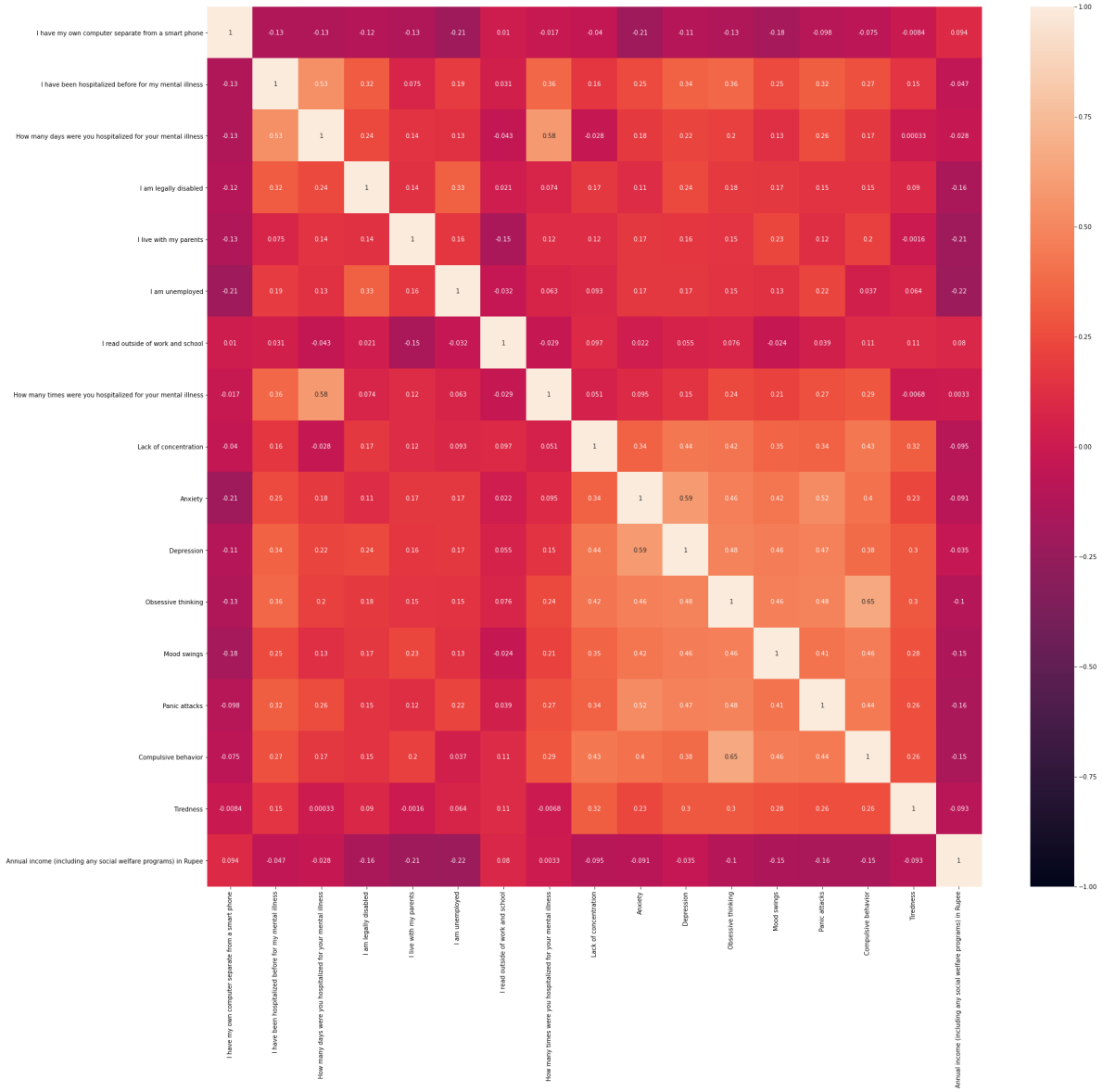
	I have my own computer separate from a smart phone	I have been hospitalized before for my mental illness	How many days were you hospitalized for your mental illness	I am legally disabled	I live with my parents	I am unemployed	I ou of sc
Annual income (including any social welfare programs) in Rupee	0.094076	-0.047485	-0.027692	-0.163601	-0.214045	-0.222513	0.08

In [18]:

```
fig = plt.figure(figsize=(22,20))
fig.add_axes([0,0,1,1])
ax = fig.get_axes()[0]
sns.heatmap(df.corr(), ax=ax, vmin=-1, vmax=1, annot=True)
```

Out[18]:

```
<matplotlib.axes._axes.Axes at 0x25b2fb707f0>
```



Label Encoding

In [19]:

```
from sklearn.preprocessing import LabelEncoder
```

In [20]:

```
education=LabelEncoder()
age=LabelEncoder()
gender=LabelEncoder()
```

In [21]:

```
df['Education']=education.fit_transform(df['Education'])
df['Age']=age.fit_transform(df['Age'])
df['Gender']=gender.fit_transform(df['Gender'])
```

In [22]:

```
df.head()
```

Out[22]:

	Education	I have my own computer separate from a smart phone	I have been hospitalized before for my mental illness	How many days were you hospitalized for your mental illness	I am legally disabled	I live with my parents	I am unemployed	I read outside of work and school	H ti ho
0	3	0	0	0.0	0	0	1	1	
1	4	1	0	0.0	0	0	0	1	
2	2	1	0	0.0	0	0	0	1	
3	5	1	0	0.0	0	1	1	1	
4	2	1	1	35.0	1	0	0	1	

Standardizing Data

In [23]:

```
from sklearn.preprocessing import StandardScaler
```

In [24]:

```
ss=StandardScaler()
```

In [25]:

```
df2=ss.fit_transform(df)
```

In [26]:

```
df2.shape
```

Out[26]:

```
(334, 20)
```

In [27]:

```
X=df.values
```

Training and Testing of Dataset

In [28]:

```
from sklearn.model_selection import train_test_split
```

In [33]:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

In [34]:

```
X_train.shape,y_train.shape
```

Out[34]:

```
((267, 20), (267,))
```

Training the Logistic regression model

In [35]:

```
from sklearn.linear_model import LogisticRegression
```

In [36]:

```
lr=LogisticRegression()
```

In [37]:

```
lr.fit(X_train,y_train)
```

```
C:\Users\amodh\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.p
y:763: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scik
it-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regre
ssion (https://scikit-learn.org/stable/modules/linear_model.html#logistic-re
gression)
  n_iter_i = _check_optimize_result(
```

Out[37]:

```
LogisticRegression()
```

In [38]:

```
y_pred=lr.predict(X_test)
```

In [39]:

```
y_pred
```

Out[39]:

```
array([0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0,
       0], dtype=int64)
```

Accuracy

In [40]:

```
from sklearn.metrics import confusion_matrix,precision_score,recall_score,accuracy_score
```

In [41]:

```
accuracy_score(y_test,y_pred)
```

Out[41]:

```
0.8208955223880597
```

In [42]:

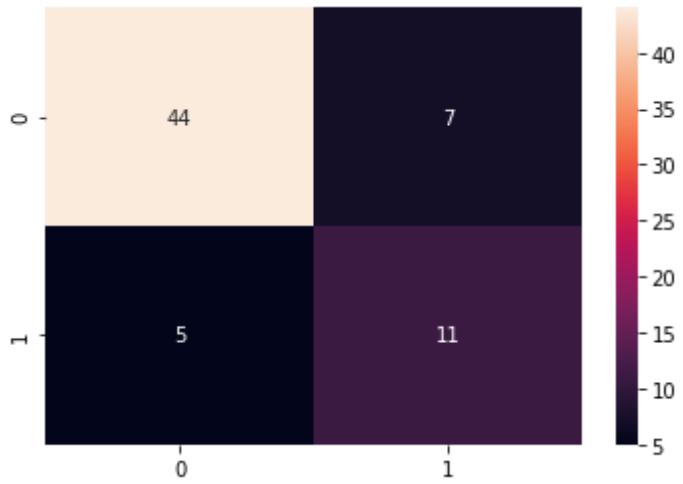
```
cf_matrix=confusion_matrix(y_test,y_pred)
```

In [43]:

```
sns.heatmap(cf_matrix, annot=True)
```

Out[43]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x25b32599f70>
```



In [44]:

```
recall_score(y_test,y_pred)
```

Out[44]:

```
0.6875
```

In [45]:

```
precision_score(y_test,y_pred)
```

Out[45]:

```
0.6111111111111112
```

In [46]:

```
from sklearn.ensemble import RandomForestClassifier
```

In [47]:

```
rf=RandomForestClassifier()
```

In [48]:

```
rf.fit(X_train,y_train)
```

Out[48]:

```
RandomForestClassifier()
```

In [49]:

```
y_pred2=rf.predict(X_test)
```

In [55]:

```
y_pred2
```

Out[55]:

```
array([0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0,
       1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0,
       0], dtype=int64)
```

In [50]:

```
accuracy_score(y_test,y_pred2)
```

Out[50]:

```
0.8955223880597015
```

In [51]:

```
recall_score(y_test,y_pred2)
```

Out[51]:

```
1.0
```

In [52]:

```
precision_score(y_test,y_pred2)
```

Out[52]:

```
0.6956521739130435
```

In [53]:

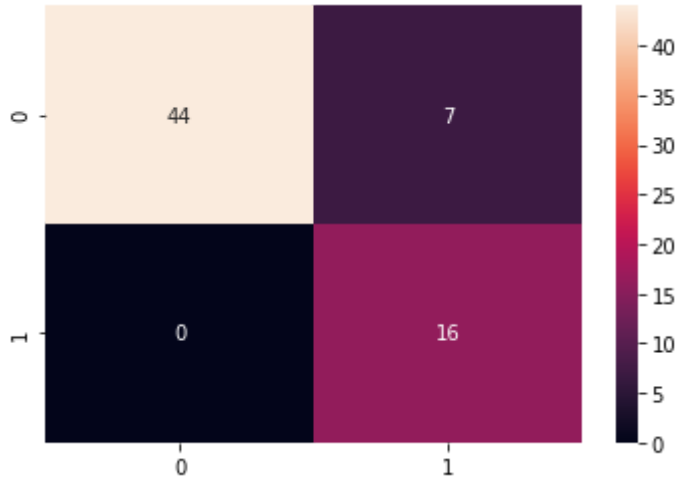
```
cf_matrix2=confusion_matrix(y_test,y_pred2)
```

In [54]:

```
sns.heatmap(cf_matrix2, annot=True)
```

Out[54]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x25b33731ca0>
```



In [ ]:

```

```