# Suicide Rates & Socioeconomic Factors

In this part we take a mental health dataset- [data](https://www.kaggle.com/datasets/bhavikjikadara/mental-health-dataset)  and growing stress prediction. This dataset comprises of various linguistic, psychological, and behavioural attributes of a very huge population. Nowadays predictive modelling is very important and using such extensive datasets we can find similarities between attributes and perform predictions on various mental health issue including suicidal rates which is directly related to growing stress in individuals. So here we use this dataset to predict the growing stress in individuals. The dataset comprises of around 300k datapoints.



This task can be viewed as a tabular classification task where we can leverage the tubular techniques like decision models. We use the labelencoder to convert the string labels into integer classes so that it can be fed into the models for training. Then we move onto to build our model for predictions. We make use of one of the most popular ensemble tree models i.e Random forest model which gives us very good results with a very high accuracy.

## 

## CODE:

## Installation

Pip install pandas sklearn

## Imports

Import pandas as pd

from sklearn.preprocessing import OneHotEncoder, LabelEncoder

from sklearn.ensemble import RandomForestClassifier

## Data Loading

df = pd.read\_csv('/content/MentalHealthDataset.csv')

X= df.drop(columns=['Growing\_Stress'],axis=1)

y= df['Growing\_Stress']

## Data processing

X = X.apply(LabelEncoder().fit\_transform)

le = LabelEncoder()

y = le.fit\_transform(y)

## Model Building

clf = RandomForestClassifier()

clf.fit(X, y)

clf.score(X,y)