

## **Problem Description**

Given Data contains details of various attributes for a population. Based on these attributes predict whether the individual belongs to low income group or high income group.

### **Files:**

Train Data

Test Data

Sample submission file

### **Instructions:**

Use Sample submission file to evaluate your model score.

Please use the exact column names and sequence of index given in sample submission file

Accuracy will be considered for evaluation

### **Dataset Description:**

- 1.age: continuous variable
2. working\_sector: sector under which the employee is working
3. financial\_weight: weighted attribute to balance the difference in the monetary and working conditions. It is a continuous variable
4. qualification: Educational qualification
5. years\_of\_education: number of years of education; continuous variable
6. tax paid: amount of tax paid by the person(continuous variable)
7. loan taken: it is a two level categorical variable defining whether the person has taken loan or no
8. marital status: categorical variable
9. occupation : area of work, a categorical variable with 14 levels
10. relationship : provides relationship status of the employee
11. ethnicity : social background (categorical variable )
12. gender: two level categorical variable
13. gain : it illustrates the financial gain of an person, it is a continuous variable
14. loss: financial loss of the person , it is a continuous variable
15. working\_hours : it is a continuous variables describing hours of work of an employee in a week
16. country: describes the origin country of an employee
17. target: Based on the given data your model will have to classify a person into high income / low income.