W.H.O Assignment

Income Group Classification Case Study

Obective:

1.analyzing the data and identify the different factors that influence the income of an individual.

2.a prediction model that can help the government formulate policies for the right pockets of the society.

3.suggestion for better progress

Task:

E D A(to read data ,

null values,

scaling issue, etc.)

-predict the correct model

-do the needful steps.

if over fitted do regularisation

Steps as follow:-

Importing needed libraries

Reading Data

Label encoding

One Hot encoding

no null values

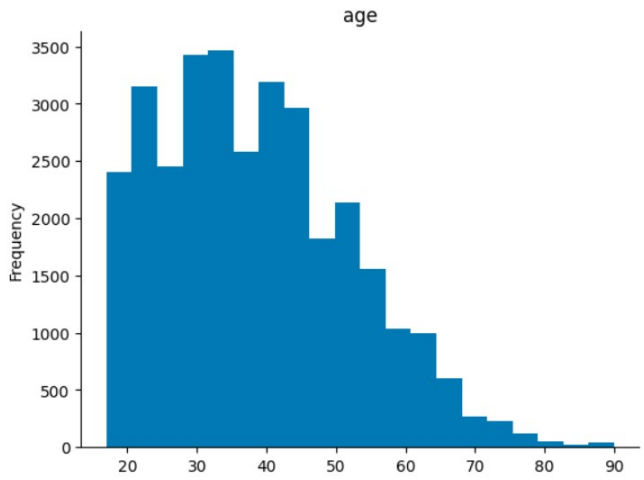
scaling issue was there normalizing data thru minmax scaler()

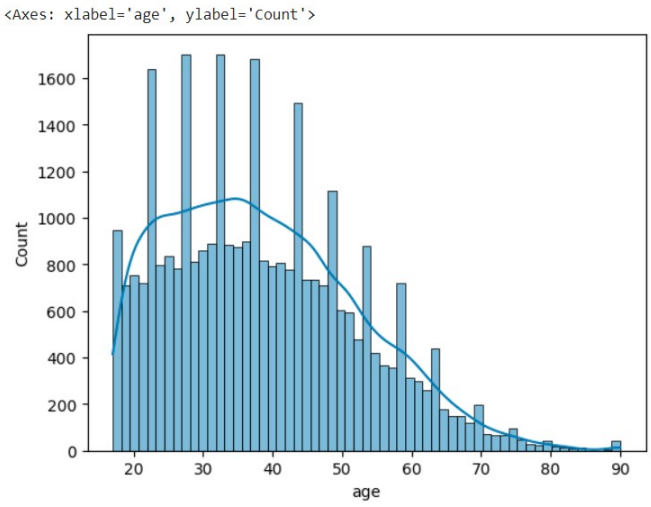
Graphical

As per my observation the main factors that influence

the income is AGE and education of years is as below:-

Age:---

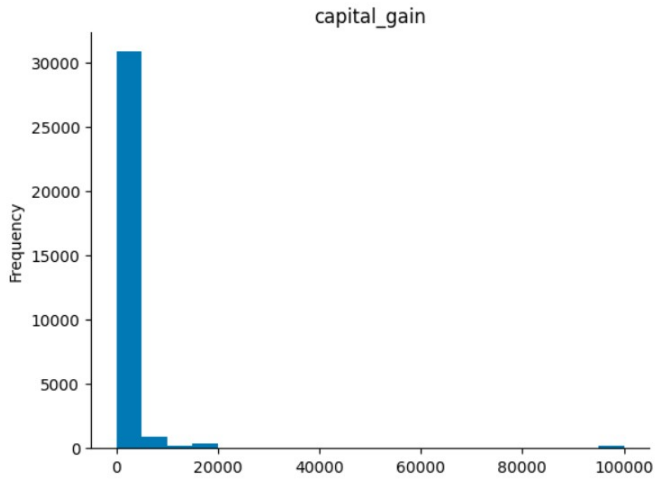




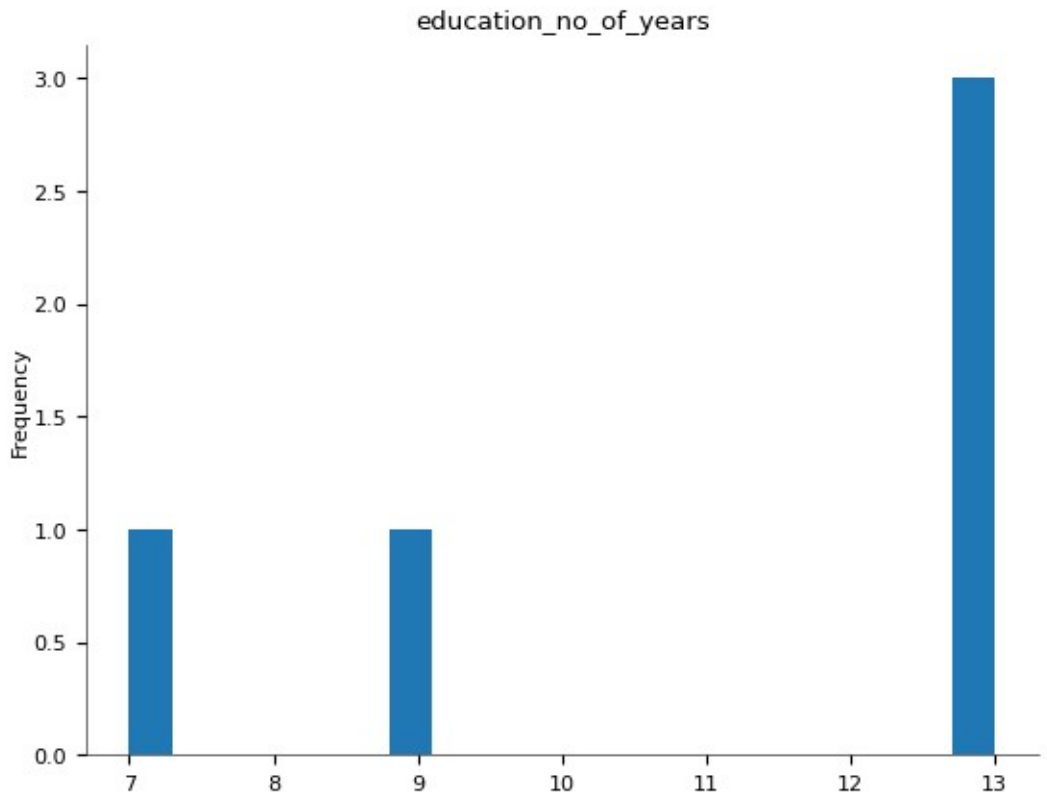
Suggesation:---

As you can see age between 25 to 45 is earning highwhile after the age of 45 income is getting low continuously so we make some most policies that benefit the age group who is above 45 or 50.

Capital Gain:-----



education no of years:---



as you can see where years of education is 13 we are getting highest income there, this is also an influancial factor that affect income so there should also be strong education some policies .

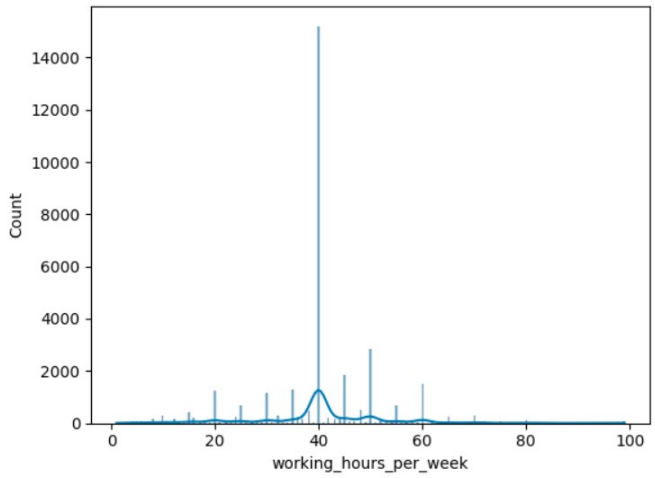
So as per my opinion I can say that the most influancial factors are

Age,

education_numbers _of_years,

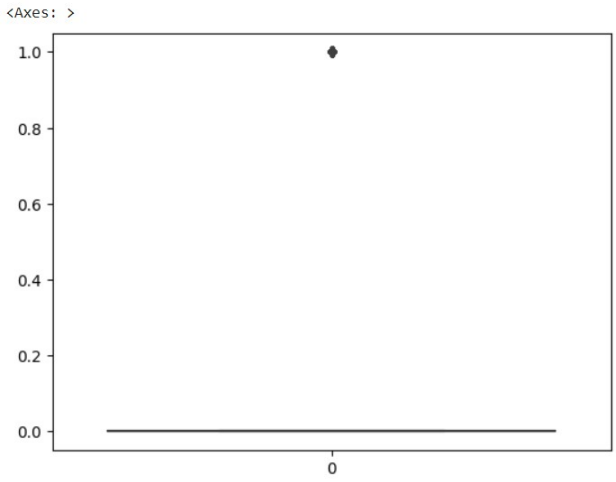
working hours per week

<Axes: xlabel='working_hours_per_week', ylabel='Count'>



I did Outlier treatment

And finally I import a machine learning model which is voting classifier after predicting I find that my model is overfitted so I did Hyper parameter tuning.



Suggestion