1) Identify your problem statement

stage 1: Machine Learning

stage 2: Supervised

stage 3: Classification

- 2) Tell basic info about the dataset (Total number of rows, columns)
  - a) Dataset contain 399 rows × 25 columns
  - b) 27 input column (age, sex, bmi, children, smoker) and 1 output column (Charges)
  - c) **Sg, rbc, pc,pcc,ba,htn,dm,cad,appet,pe and ane** are categorical (Ordinal) column so we have to convert as numerical data
- 3) Mention the pre-processing method if you're doing any (like converting string to number nominal data)

# **Dataset for Before Preprocessing**

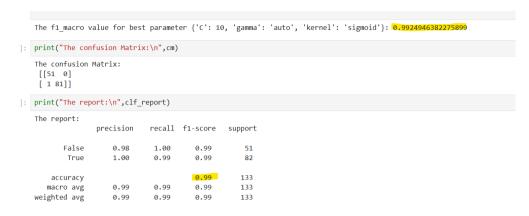


# **Dataset for After Preprocessing**

a) Converting string to number - Ordinal Mapping (Label Encoder)

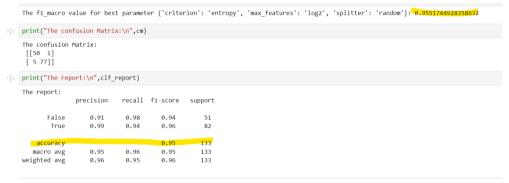


- 4. Develop a good model with good evaluation metric. You can use any machine learning algorithm; you can create many models. Finally, you have to come up with final model..
  - a) SVM Grid Search



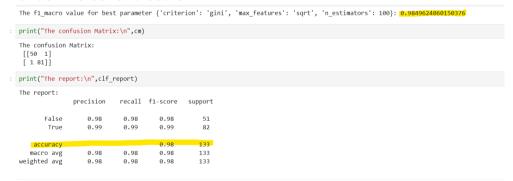
Accuracy is 0.9924946382275899

## b) Descension tree Grid Search



### Accuracy is 0.812447479

#### c) Random Forest Grid Search



Accuracy is 0.9849624060150376

### **Final Result**

SVM Grid Search is the best model which gives high Accuracy is 0.9924946382275899 compare to other model