**CLASSIFICATION ASSIGNMENT**

**Problem statement**

The client is from a health sector and they want help in classifying the people based on the results from their health check reports.

**Dataset**

The dataset contains 400 rows and 25 columns of data.

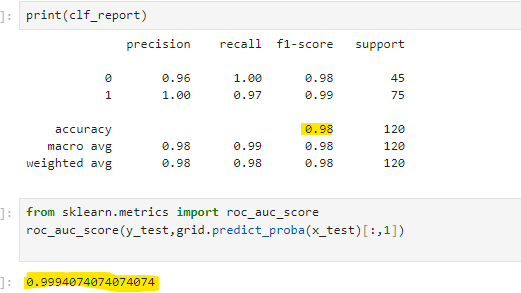
**Data pre-processing**

The data is a mix of categorical and numerical value. Here the categorical data is converted into numerical data using get dummies function from pandas.

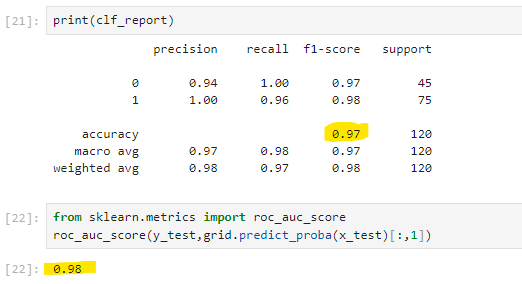
**Creating a good model**

The goal is to create a model with good accuracy and roc\_auc\_curve using machine learning algorithm. Here the output is categorical so, “classifier” is used. It’s a supervised learning as the input and outputs is clearly defined.

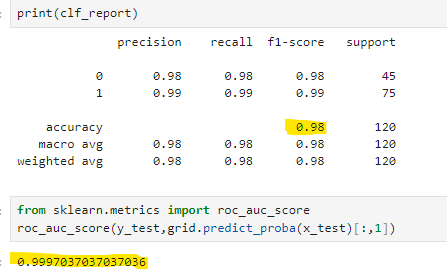
**1.SUPPORT VECTOR MACHINE**



**2.DECISION TREE**

****

**3.Random Forest**

****

**Final Model**

The final model is **Machine learning>>Classification>>Random Forest**

Justification: Even the accuracy is same for SVM and Random Forest the roc\_auc\_curve value of Random Forest is high comparing to other algorithms.