# **Classification Assignment**

**Problem Statement or Requirement:**

A requirement from the Hospital, Management asked us to create a predictive model which will predict the Chronic Kidney Disease (CKD) based on the several parameters. The Client has provided the dataset of the same.

* **Identify your problem statement**

Here the problem statement is to identify whether any individual/patient will have CKD in near future based on several i/p parameters that the hospital has captured. Since the i/p is in numerical formal we can use **MACHINE LEARNING** for providing the solution. Also, since the requirement is clear and we have both the i/p and o/p data handy this will come under **SUPERVISED LEARNING**. Further the o/p is categorical and hence we would go ahead with **CLASSIFICATION**

* **Tell basic info about the dataset (Total number of rows, columns)**

The dataset that has been provided:

* + 1. Column🡪 25
       1. 13- numerical
       2. 12- categorical
    2. Rows🡪 399
* **Mention the pre-processing method if you’re doing any (like converting string to number – nominal data)**

Since the data is NOMINAL, we should be using **ONE-HOT ENCODING** to update to numerical format for our Python code to handle it

* **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.**

Used GRIDSERACH for all the 6 algorithms below:

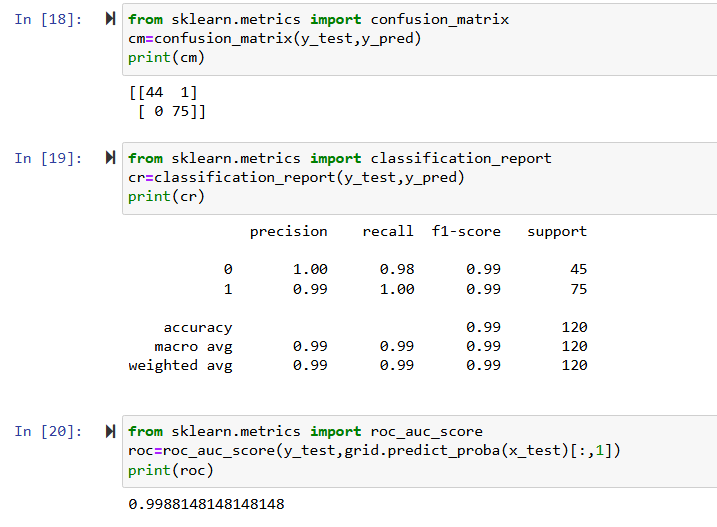
* Logistic Regression
* SVC
* Decision Tree
* Random forest
* KNN
* Naïve Bayes
* **All the research values of each algorithm should be documented. (You can make tabulation or screenshot of the results.)**

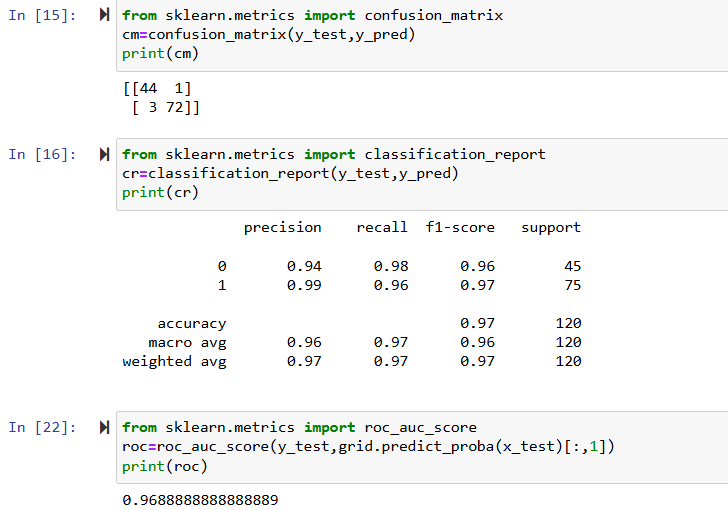
1. **Logistic Regression**

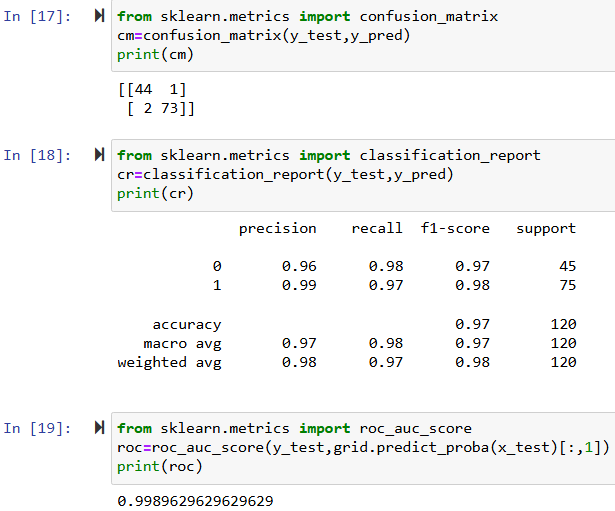
A screenshot of a computer

AI-generated content may be incorrect.

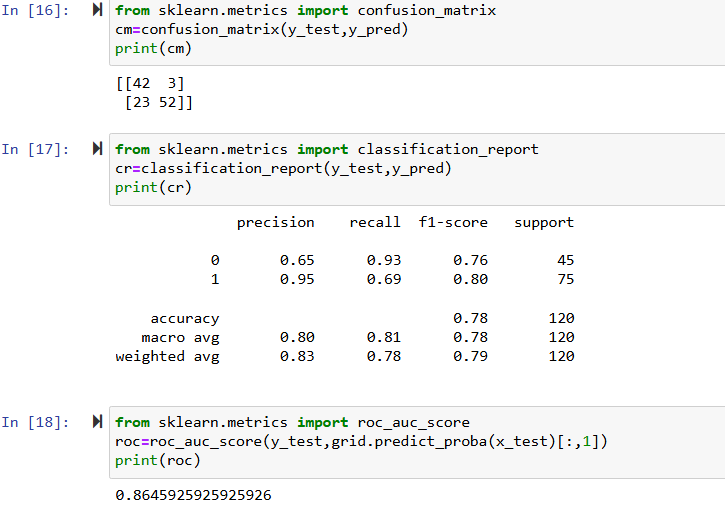
1. **SVC**

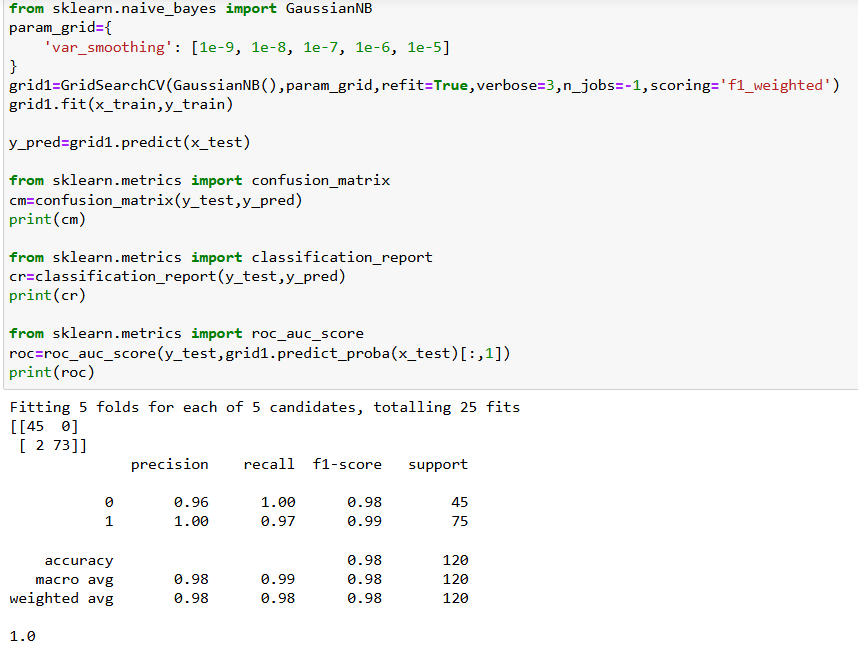


1. **DecisionTree **
2. **Random forest**



1. **KNN**



1. **NaïveBayes** 

* **Mention your final model, justify why u have chosen the same.**

The final chosen model would be SVC with ACCURACY = 0.99% and ROC=0.998%

