**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.

1. Identify your problem statement

Need to predict CKD – Yes / No

1. Data table given – so we can choose ML
2. Both Input and Output are clearly given – Supervised Learning
3. Output is categorical – Classification
4. Tell basic info about the dataset (Total number of rows, columns)

399 rows × 25 columns

Input Contains both numerical and categorical dataset

Output contains categorical dataset

Index(['age', 'bp', 'sg', 'al', 'su', 'rbc', 'pc', 'pcc', 'ba', 'bgr', 'bu',

'sc', 'sod', 'pot', 'hrmo', 'pcv', 'wc', 'rc', 'htn', 'dm', 'cad',

'appet', 'pe', 'ane', 'classification']

1. Mention the pre-processing method if you’re doing any (like converting string to number – nominal data)

Using get dummies function convert the str to flt

1. Develop a good model with good evaluation metric. You can use any

machine learning algorithm; you can create many models. Finally, you

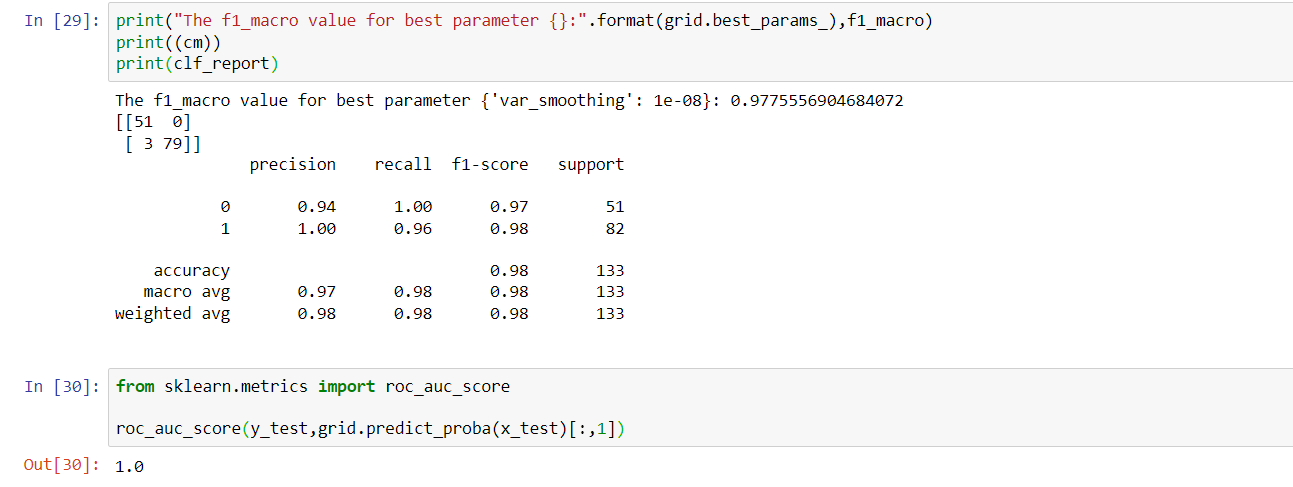
have to come up with final model.

Created models in Logistics regression, KNN , Navie bayes and SVM

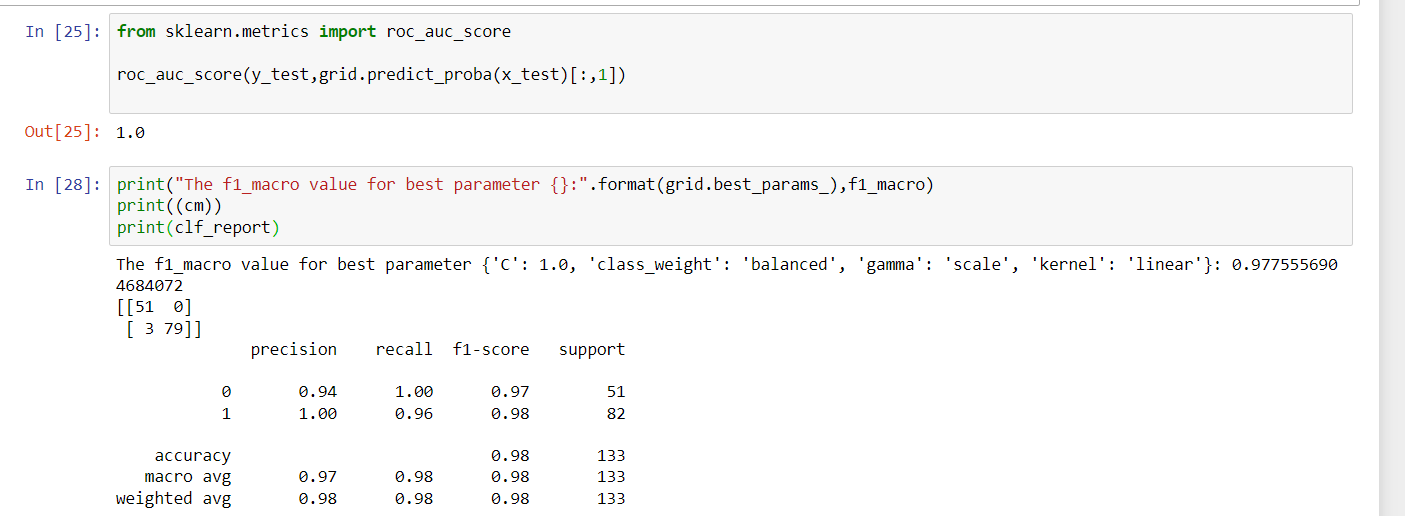
* roc\_auc\_score in KNN - 0.9878048780487805
* roc\_auc\_score in Log – 0.5
* roc\_auc\_score in NBs
  + Gaussian NB’s - 1.0
  + Complement NB’s – -ve inputin data, NAfor C NB
* roc\_auc\_score in SVM – 1.0

1. All the research values of each algorithm should be documented.

**Gaussian NB’s**

****

**SVM:**

****

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

We can choose either SVM or Gaussian Navie Bayes. Since both are having giving same results.