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.

DATASET:[D:\Hope AI\Machine Learning\Classification\Social\_Network\_Ads.csv](file:///D:\Hope%20AI\Machine%20Learning\Classification\Social_Network_Ads.csv)

1. Identify your problem statement

*My client’s requirement is to predict the Chroni Kidney Disease(CDK) using some parameters of the patients. So, This is typically classification problem.*

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

*399 rows × 25 columns*

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

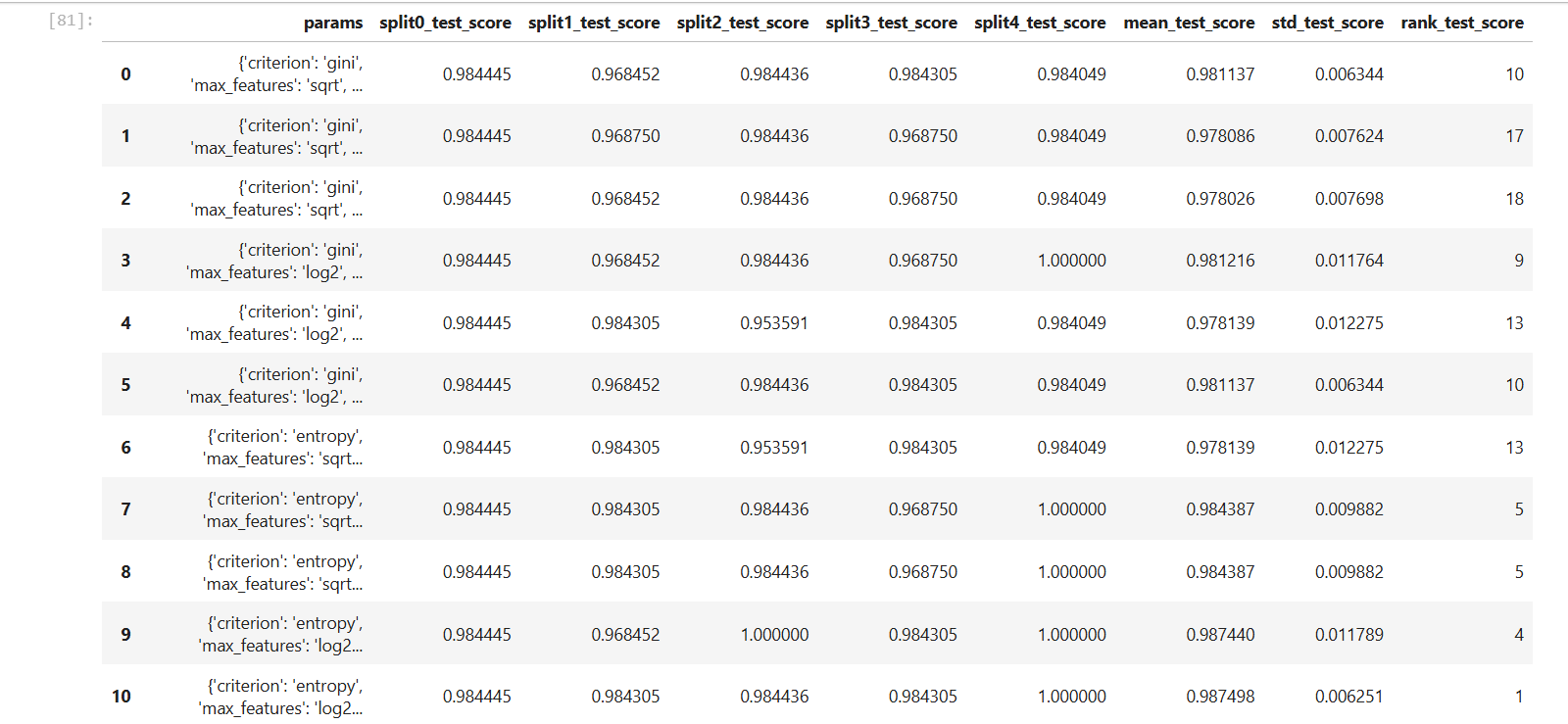
=>*I have done preprocessing of data in the columns ” sg, rbc, pc, pcc, ba, htn, dm, cad, appet, pe, ane, classification”.For preprocessing I have use get\_dummies in-built function of panda library, which convert the categorical data into numerical data.*

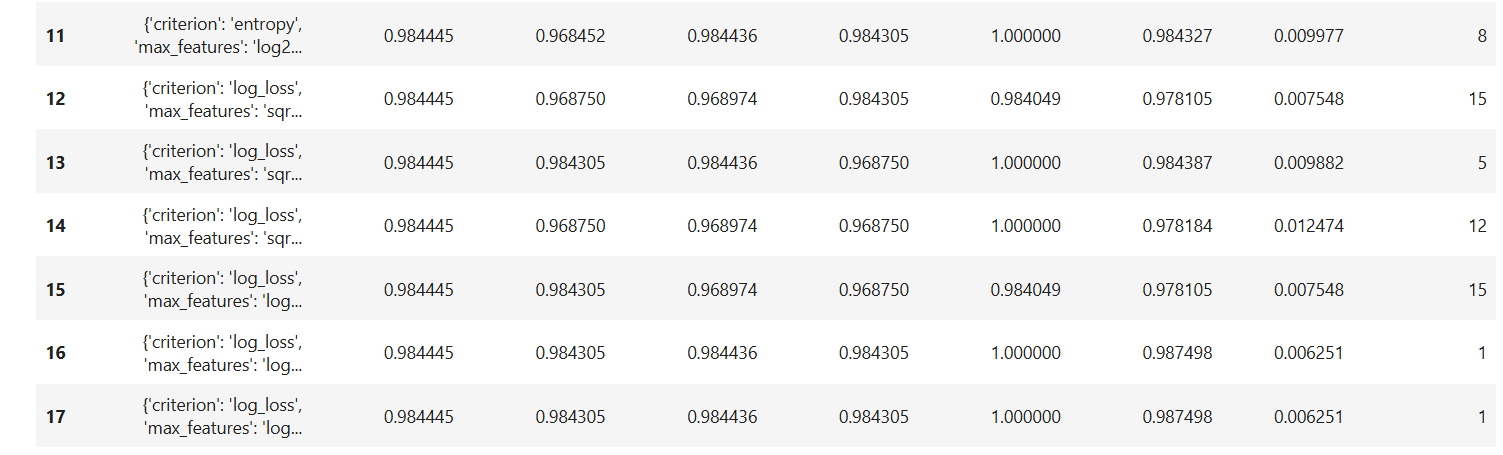
*=> I have also detect that some columns like rc,, wc have some more outliers than others using boxplot function of seaborn library. After detecting the outliers, I remove it using the method called percentile method specically I have use capping instead of trimming to prevent data loss.*

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.

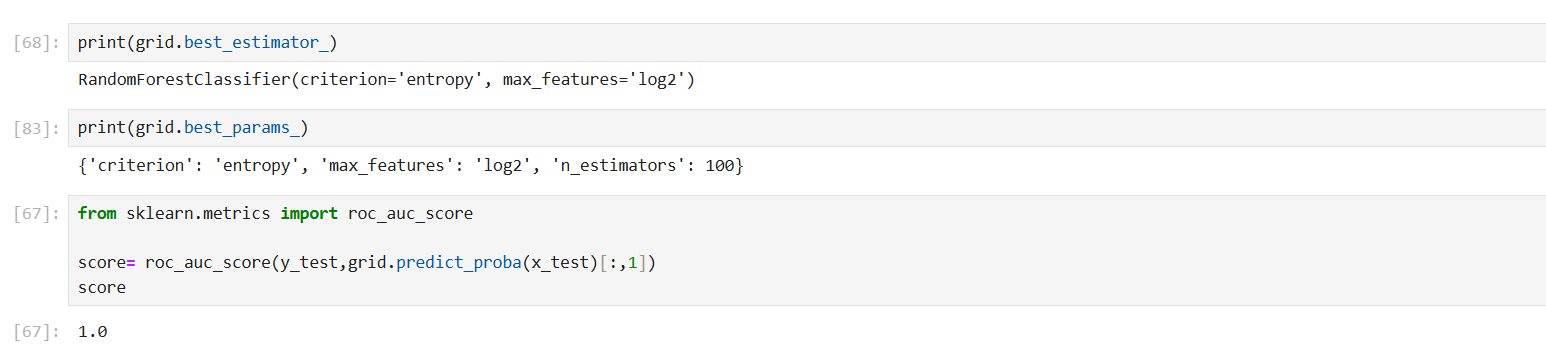
*I have developed models using different algorithm that suits well for classification problem.*

5.) All the research values of each algorithm should be documented. (You can make tabulation or screenshot of the results.)

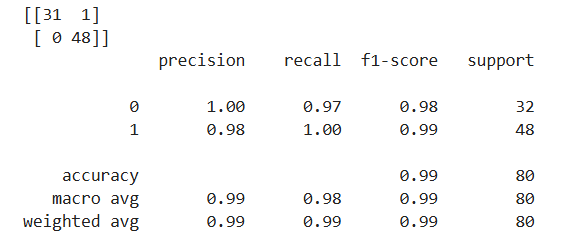
RANDOM FOREST:

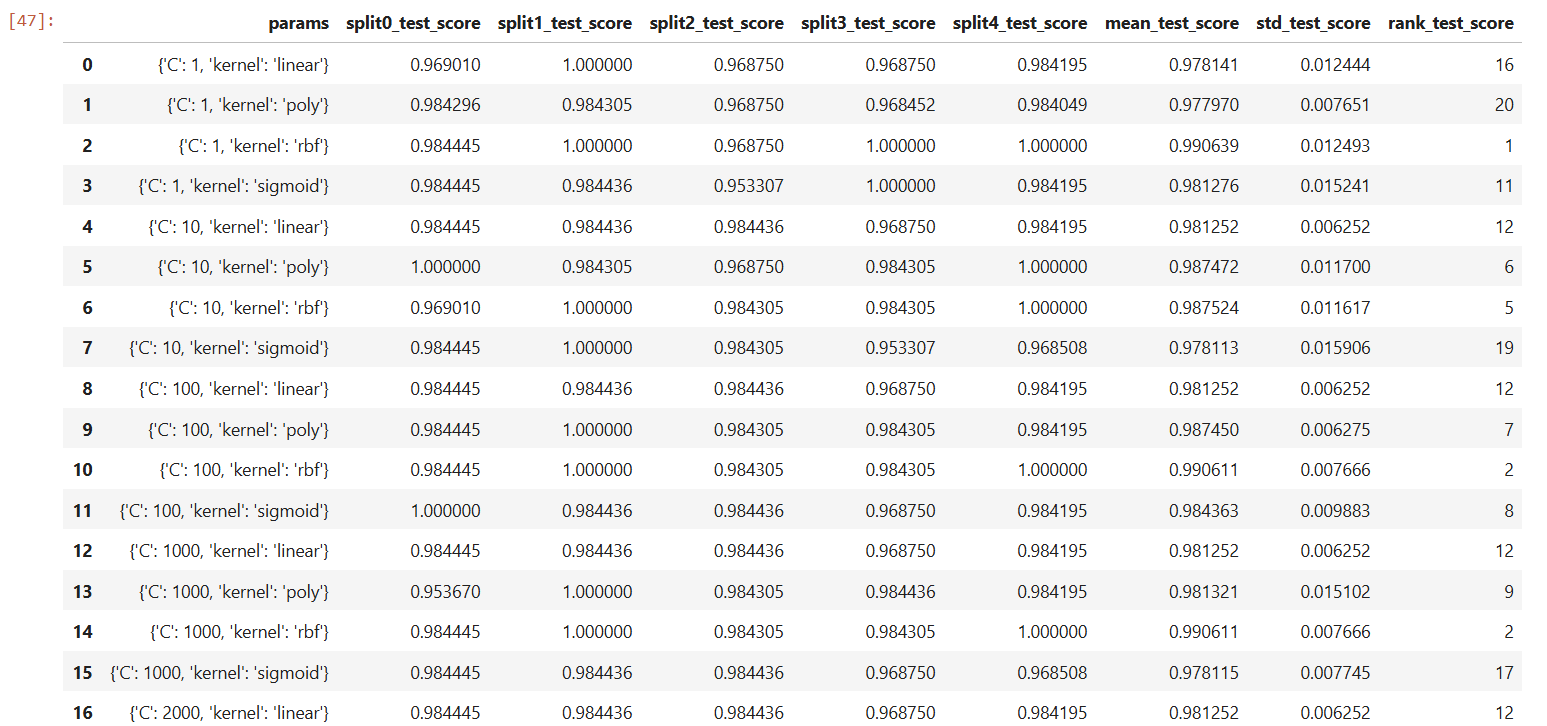
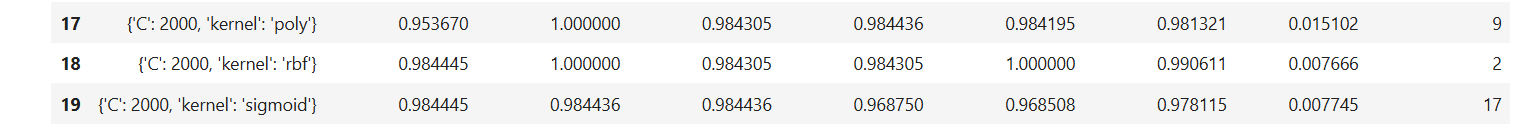


Best Parameters:

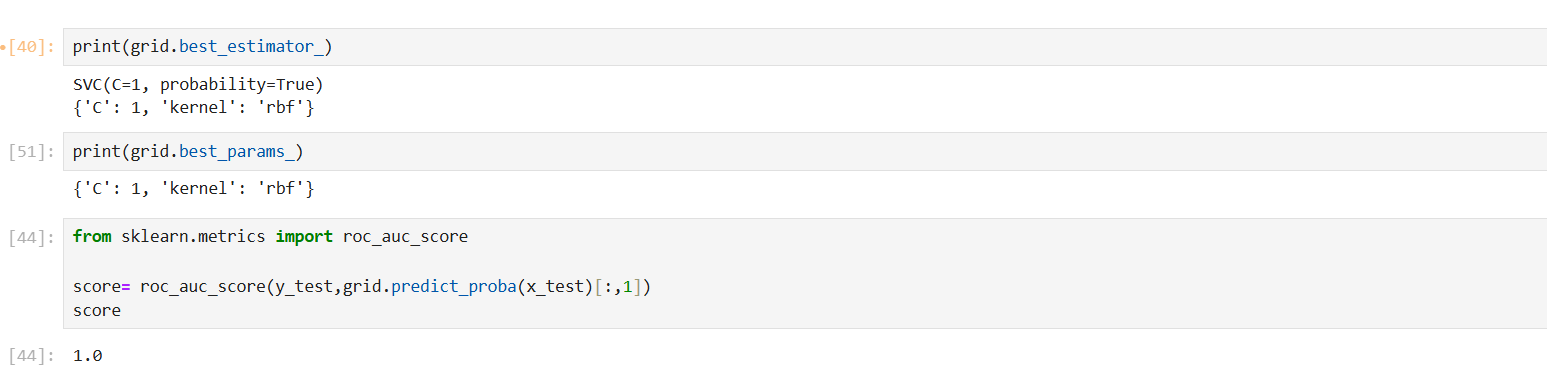


Confusion Matrix:

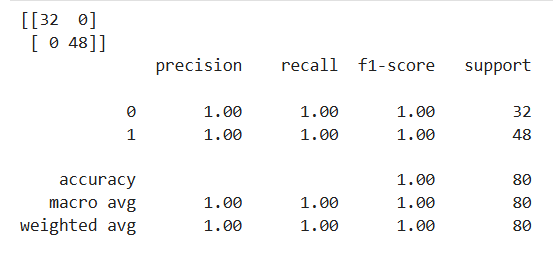


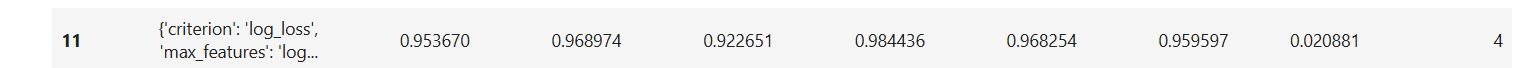
Support vector machine:

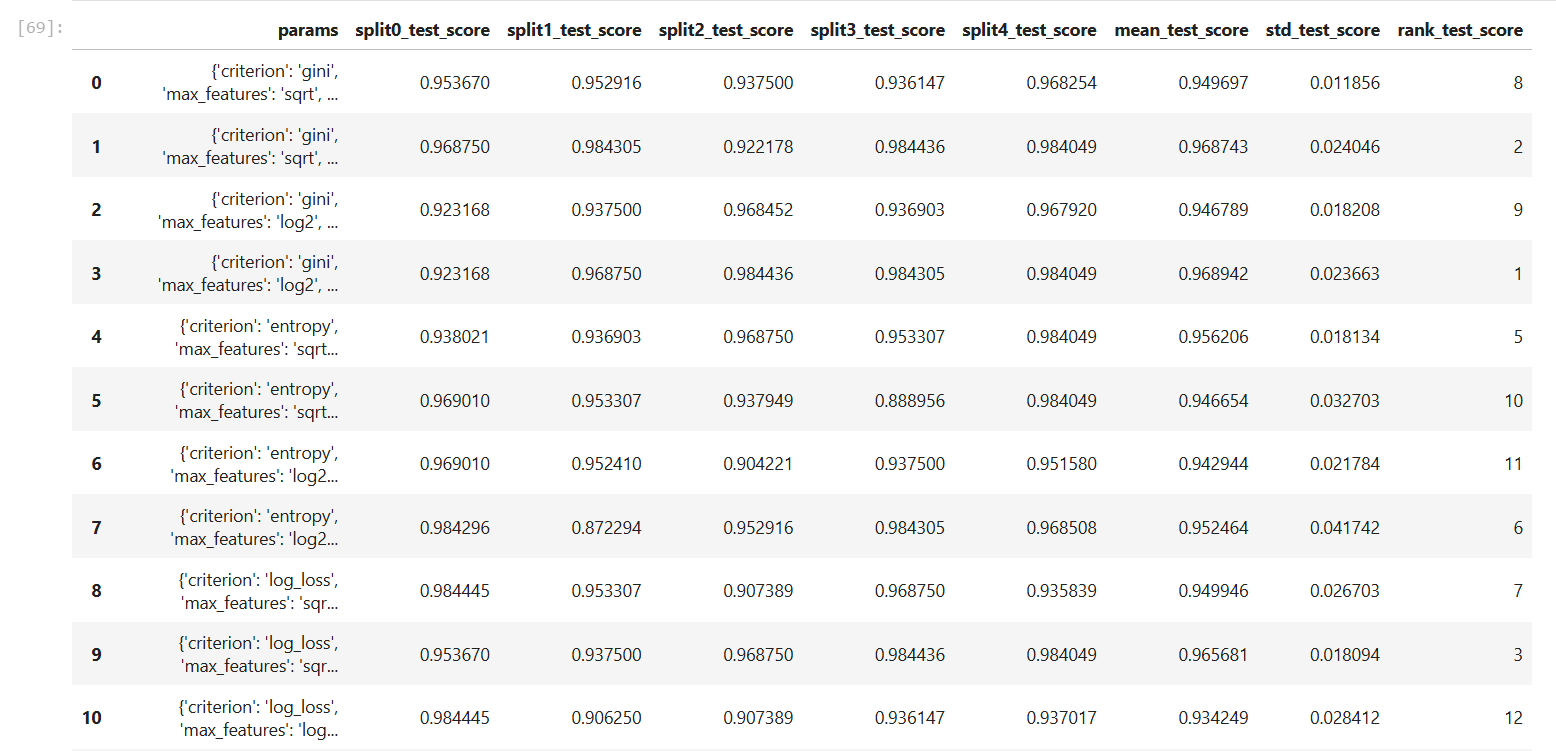
Best parameters:



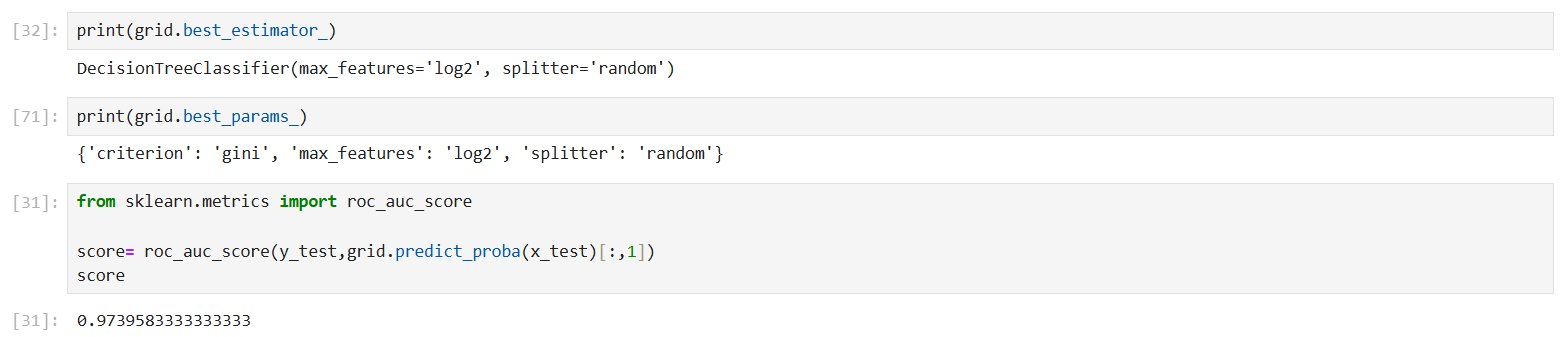
Confusion Matrix:



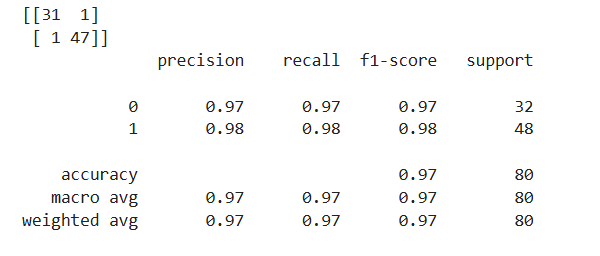
DECISION TREE:

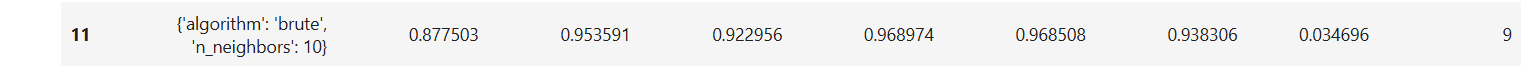


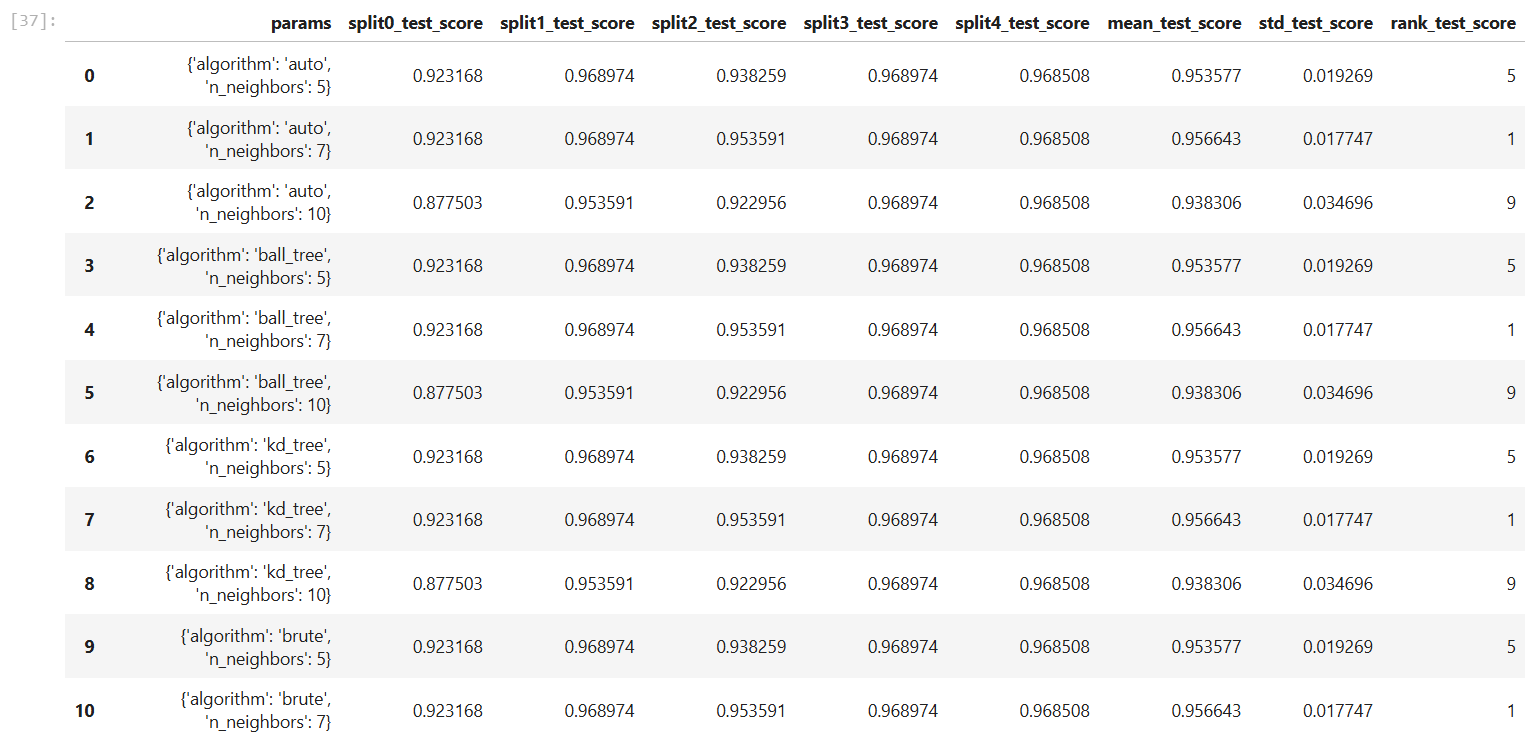
Best parameters:



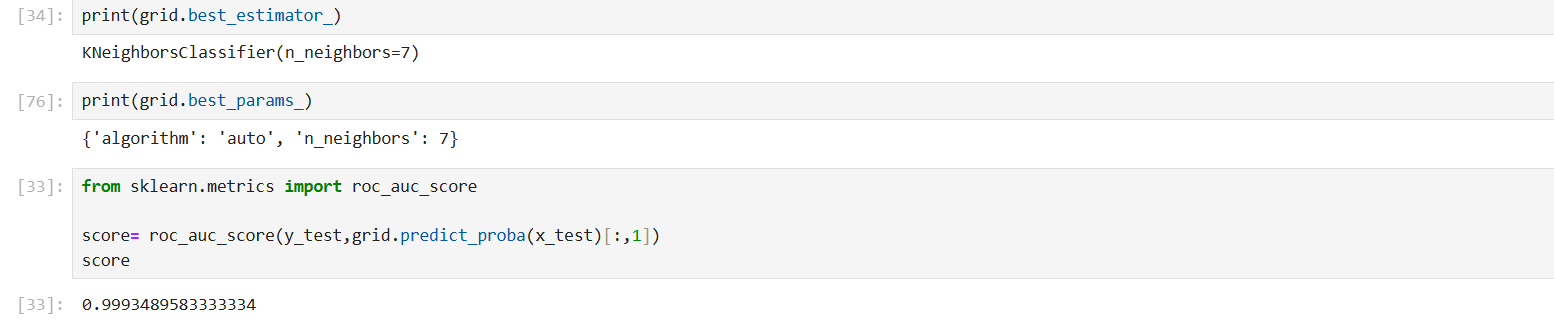
Confusion matrix:



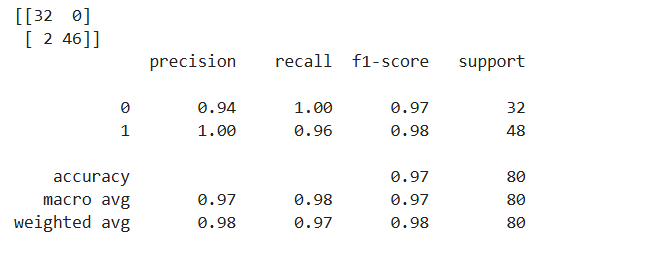
K Nearest Neighbours:



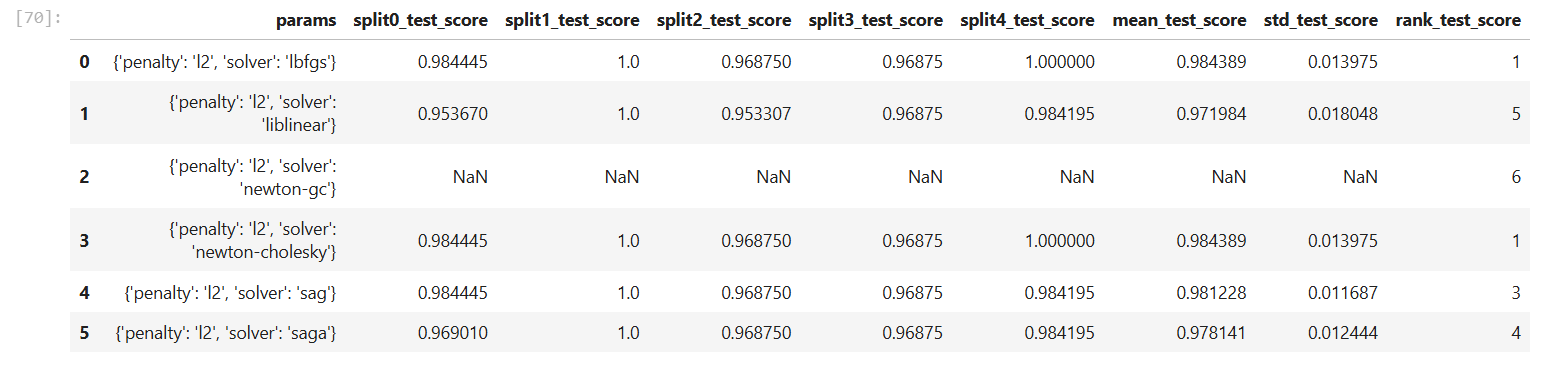
Best parameters:

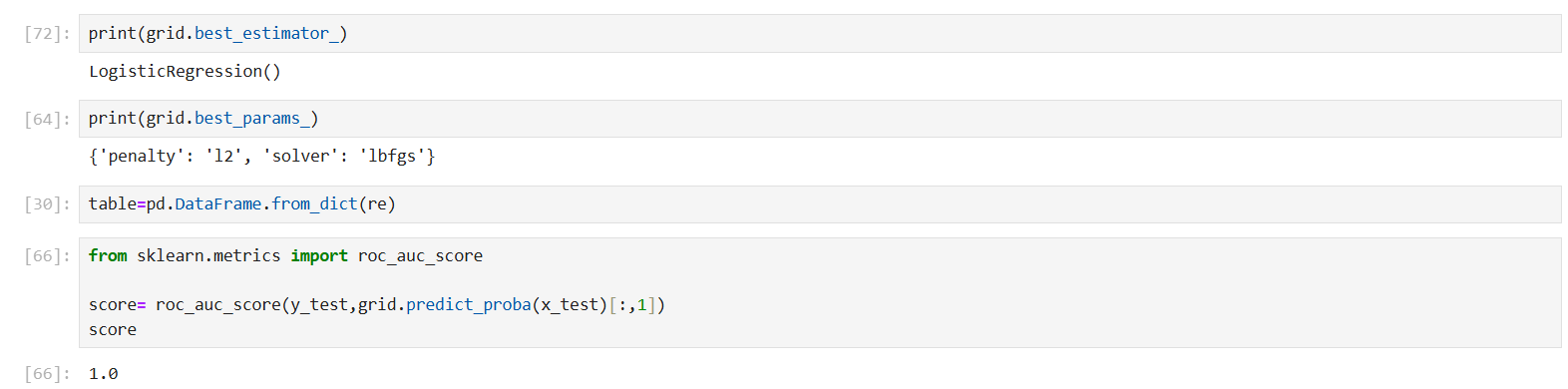


Confusion matrix:

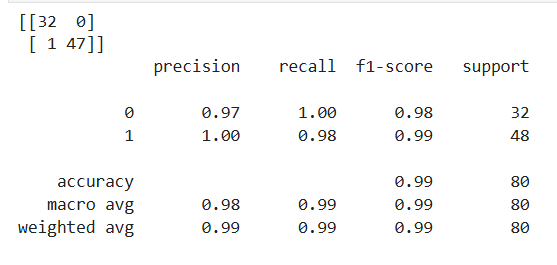


LOGISTIC REGRESSION:

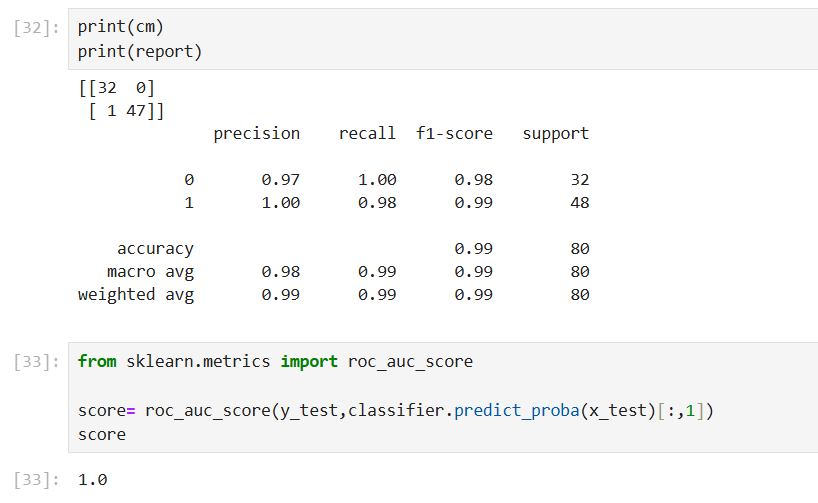


Best parameters:

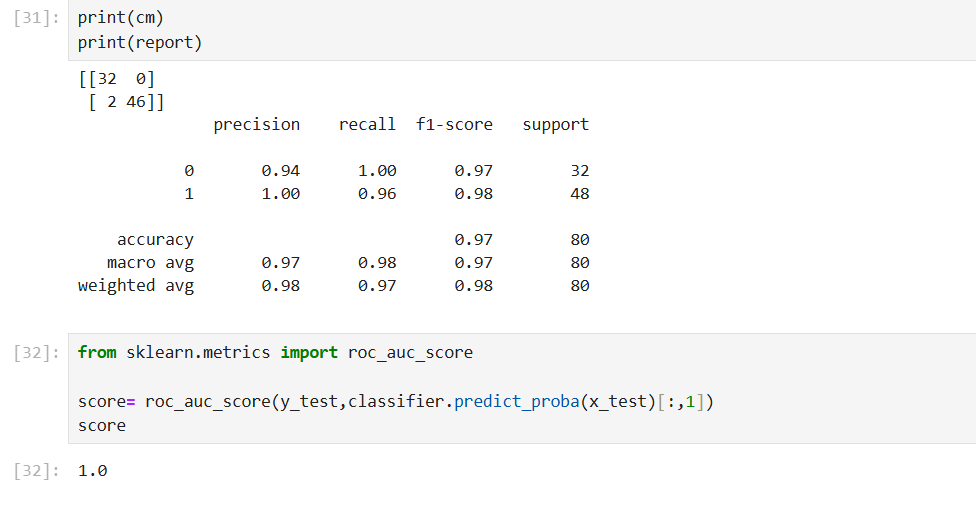
Confusion matrix:



Gaussian Naïve bayes:



bernoulli naïve bayes:



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

*I have chosen support vector machine model SCV with parameter c:1 and kernel:rbf*

*REASON:*

*=> First, it has the accuracy of one*

*=> Second, it has high recall and high precision*

*=> Third, it also has less Type 1 and Type 2 error*