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

Hospital management has provided the details about the patients health inspections and asked us to predict the possibilities of Kidney diseases using the previous history

Below are the details about dataset and modelling.

1. Machine learning works because the input is numerical
2. This comes under supervised learning, the input and the output is defined. And the data is complete.
3. Classification model- the expected output is true or false

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

399 rows and 28 columns.

3.) Mention the pre-processing method if you’re doing any (like converting

string to number – nominal data)

pd.get\_dummies(dataset,drop\_first = True) is used to convert the ordinal data to nominal data

rbc,pc parameters are normal and abnormal

pcc,ba parameters are present and not present

htn,dm,cad,appet,pe,ane are yes/no params

sg is categorized as a,b,c,d

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.

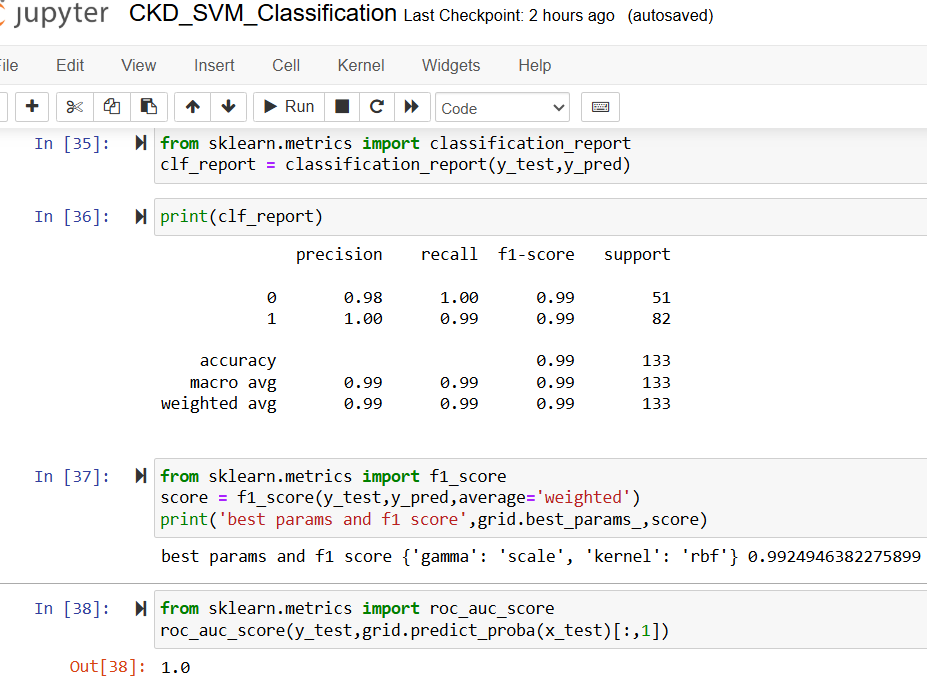
The best params is below and the model is SVM

best params and f1 score {'gamma': 'scale', 'kernel': 'rbf'} 0.9924946382275899

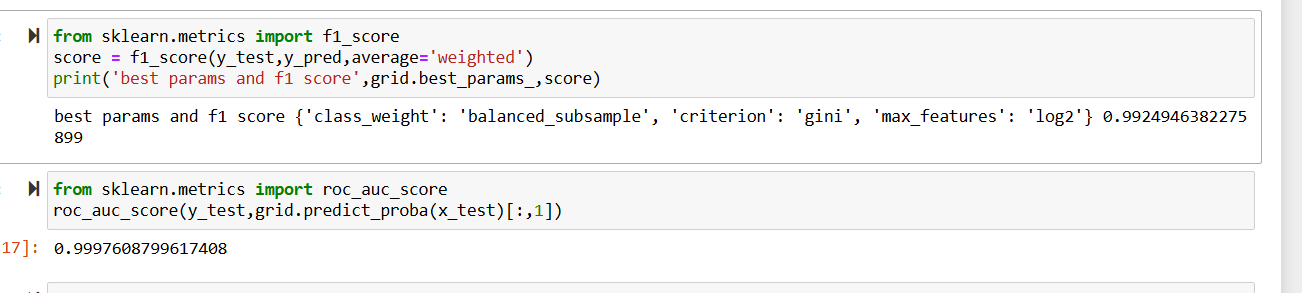
Roc\_auc\_score =1

5.) All the research values of each algorithm should be documented. (You

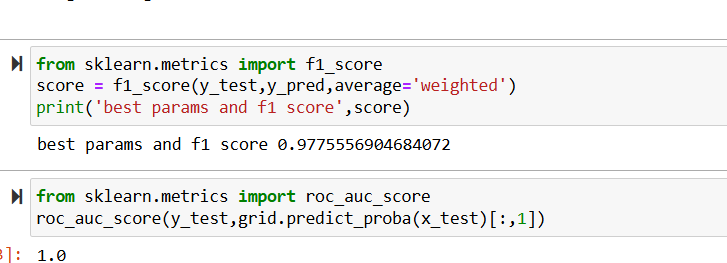
can make tabulation or screenshot of the results.)



Random forest



Naïve Bayes



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

The best model is chosen as SVM because the Roc score is 1 and F1 score is 0.99