# 1.) Identify your problem statement

Supervised Learning

Machine Learning

classification

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

RangeIndex: 399 entries, 0 to 398

Data columns (total 25 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 age 399 non-null float64

1 bp 399 non-null float64

2 sg 399 non-null object

3 al 399 non-null float64

4 su 399 non-null float64

5 rbc 399 non-null object

6 pc 399 non-null object

7 pcc 399 non-null object

8 ba 399 non-null object

9 bgr 399 non-null float64

10 bu 399 non-null float64

11 sc 399 non-null float64

12 sod 399 non-null float64

13 pot 399 non-null float64

14 hrmo 399 non-null float64

15 pcv 399 non-null float64

16 wc 399 non-null float64

17 rc 399 non-null float64

18 htn 399 non-null object

19 dm 399 non-null object

20 cad 399 non-null object

21 appet 399 non-null object

22 pe 399 non-null object

23 ane 399 non-null object

24 classification 399 non-null object

dtypes: float64(13), object(12)

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

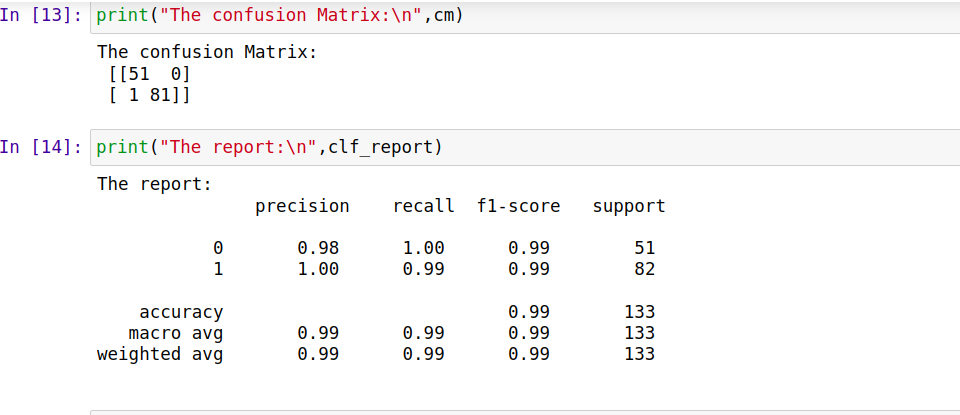
# Yes i converted column rbc,pc,pcc,ba,htn,dm,cad,appet,pe,ane classificationstring column to number

# 5.) All the research values (roc\_score of the models) should be documented.

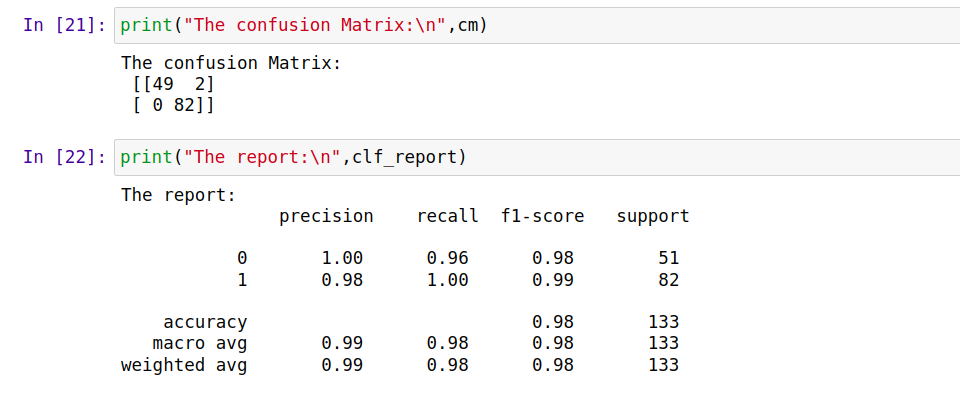
# (You can make tabulation or screenshot of the results.)

Hyper Tuning parameter

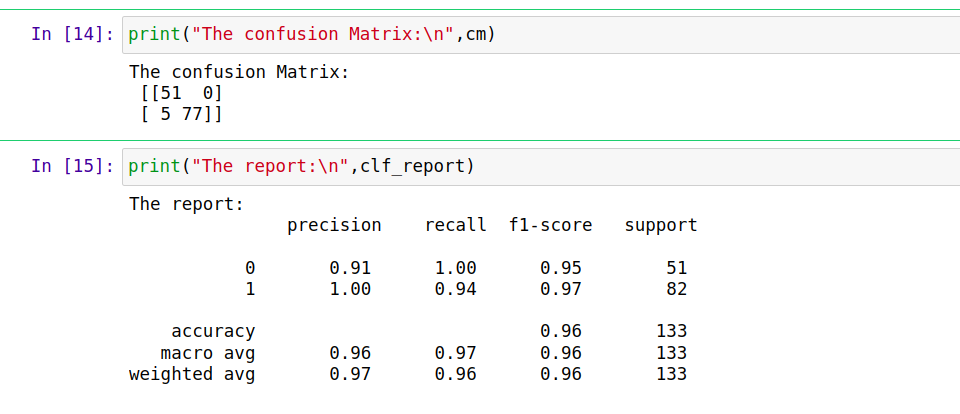
Support vector machine == 99



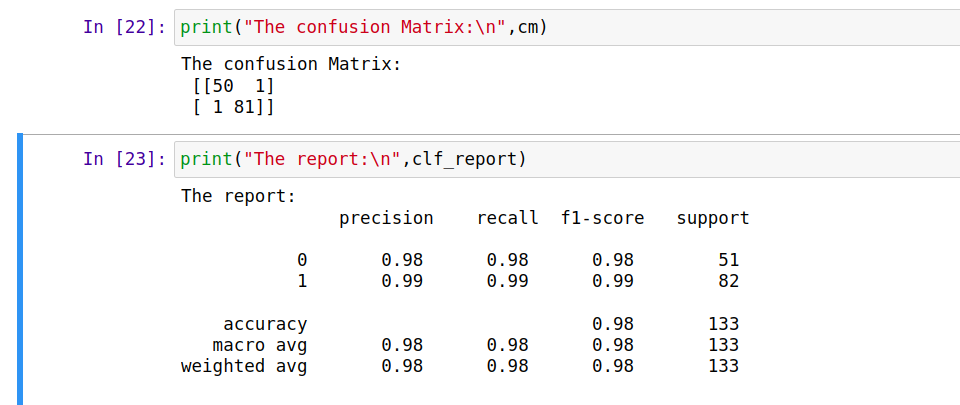
Logistic classification == 98



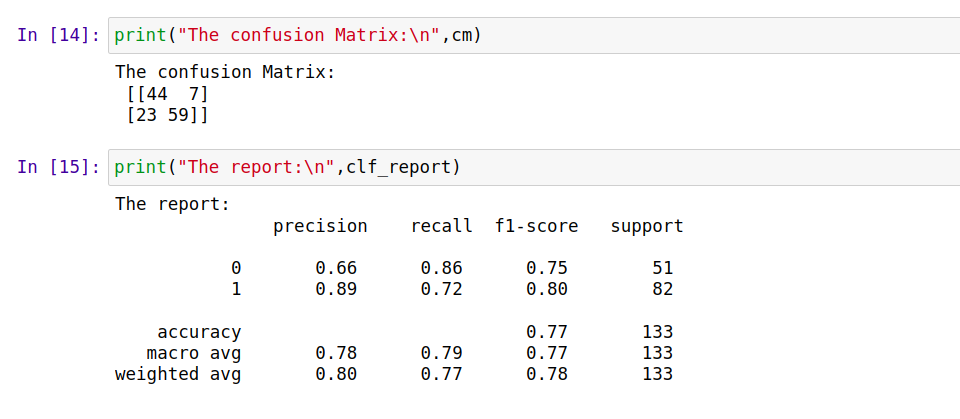
Decision tree ==96



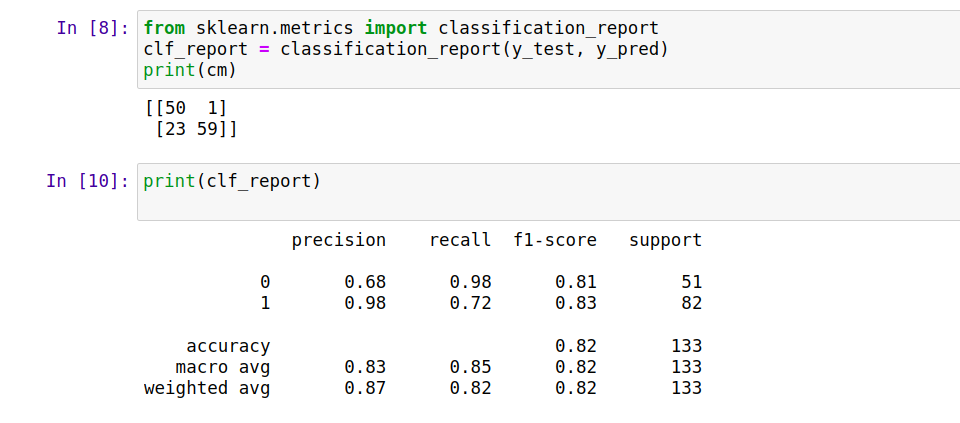
Randomforest == 98



# KNN == 77



# Naive == 82



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

i suggest support vector Machine because i got more Roc value 0.99