Assignment –Classification Algorithm

1. Identify your problem statement:

Stage 1: Machine Learning(inputs are present in numerical)

Stage 2: Supervised Learning(input and outputs are clear)

Stage 3: Supervised -Classification(output is categorical value)

2. Tell the basic info about your dataset:

There are 399 rows × 25 columns in the given dataset.

Input columns/dependent: 'age', 'al', 'su', 'bgr', 'sc', 'pot', 'hrmo', 'pcv',

'wc', 'rc', 'sg\_b', 'sg\_c', 'sg\_d', 'sg\_e', 'rbc\_normal', 'pc\_normal',

'pcc\_present', 'ba\_present', 'dm\_yes'.

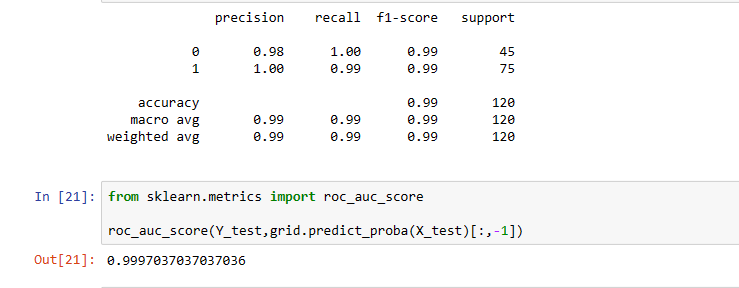
Output column/independent: ‘classification\_yes’.

3. Mention the pre-processing method:

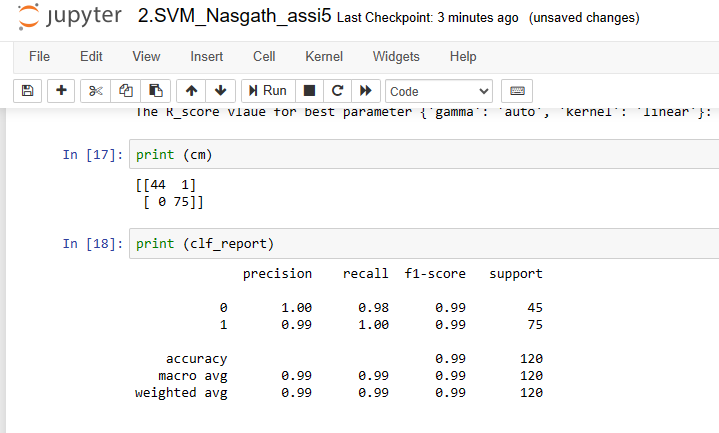
From the given dataset some columns are string format,so converting to nominal data(one hot encoding) is used in pre-processing method.

4. Develop a good model with Evaluation Matrix values:

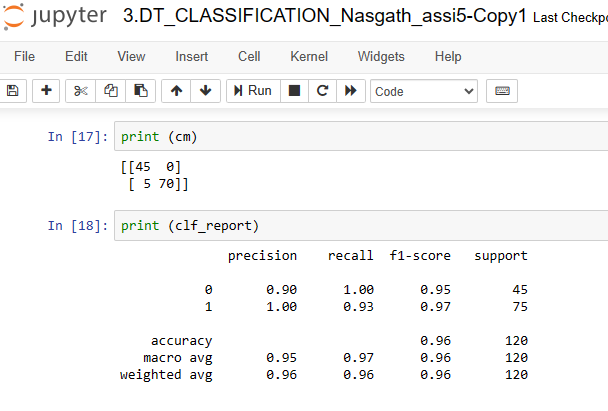
* **LOGISTIC REGRESSION: (Evaluation Matrix value) = 0.99**

****

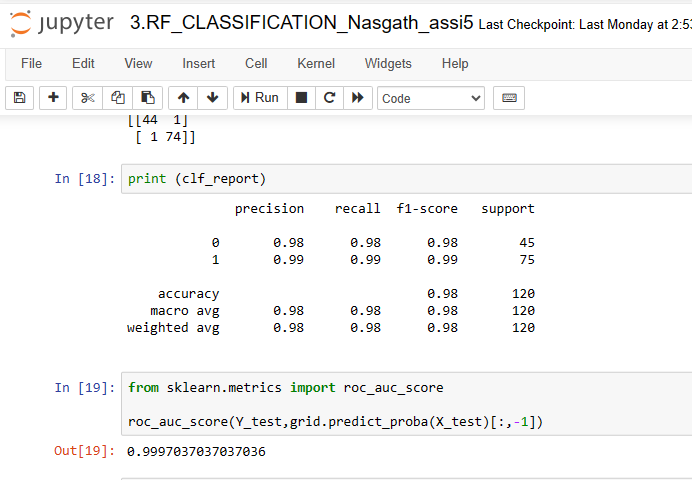
* **SVM**



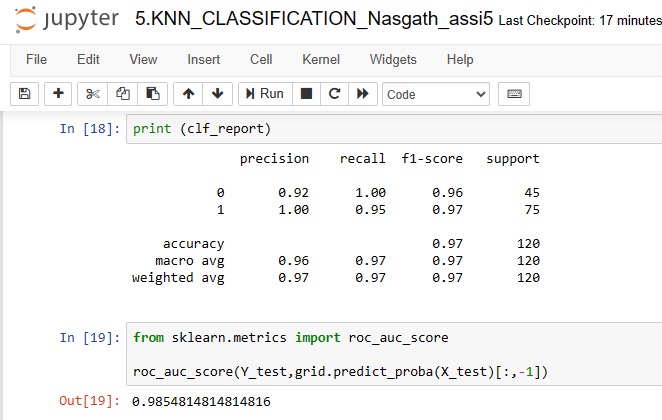
* **DECISION TREE**

****

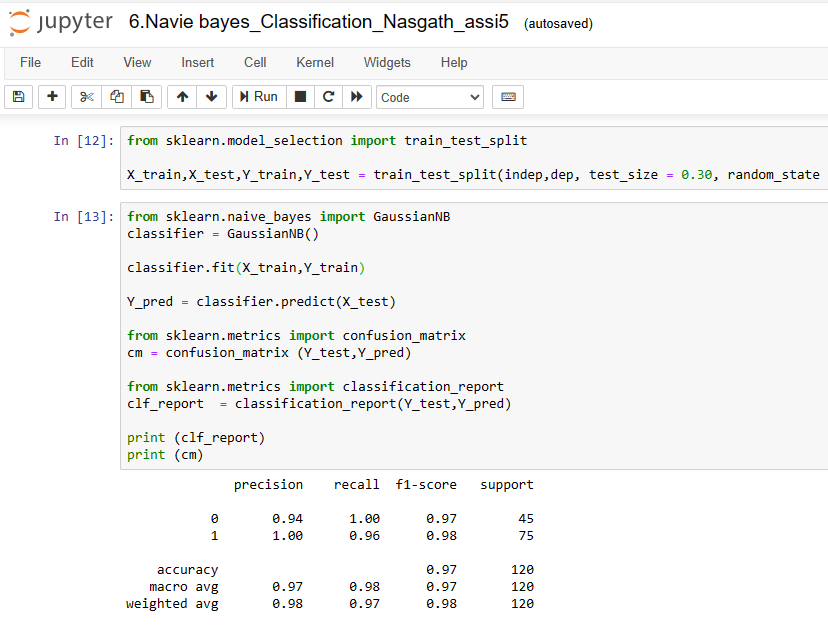
* **RANDOM FOREST:**



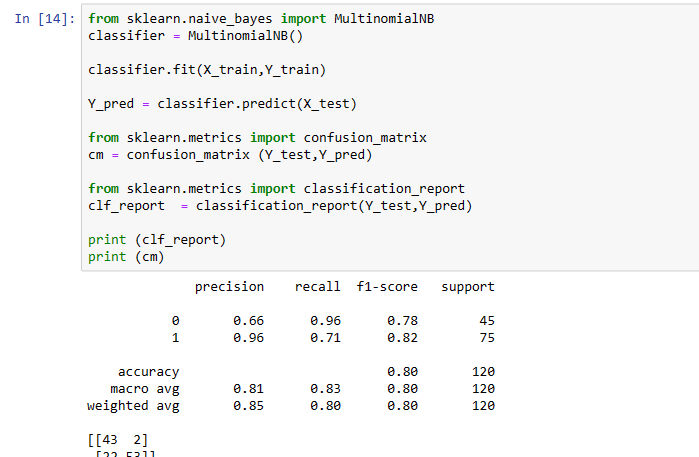
* **KNN**

****

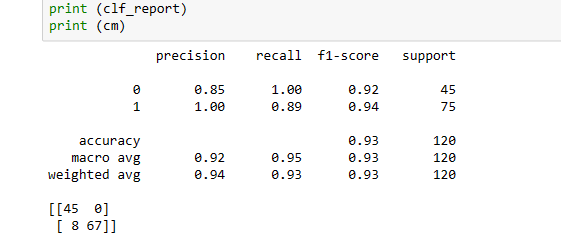
* **Navie Bayes: Gaussian NB**

****

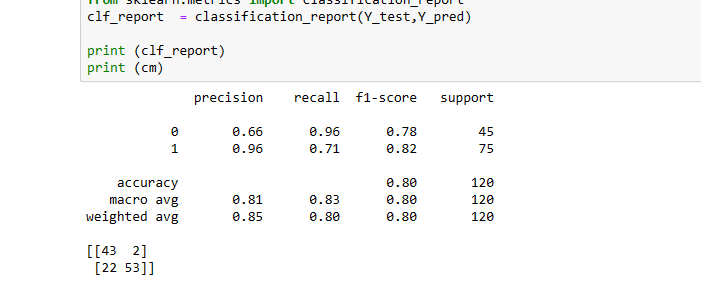
* **Multinomial NB:**

****

* **BernoulliNB:**

****

* **ComplementNB:**

****

6. Mention the final model and justiy:

The final model I have chosen is Logistic regression and SUPPORT VECTOR MACHINE, since the accuracy value **0.99 which is nearly to 1**.