Hope Artificial Intelligence

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

Stage 1- Domain- Machine Learning

Stage 2- Learning- Supervised Learning

Stage 3- Classification

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

No of Rows: 399
No of Columns: 28

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

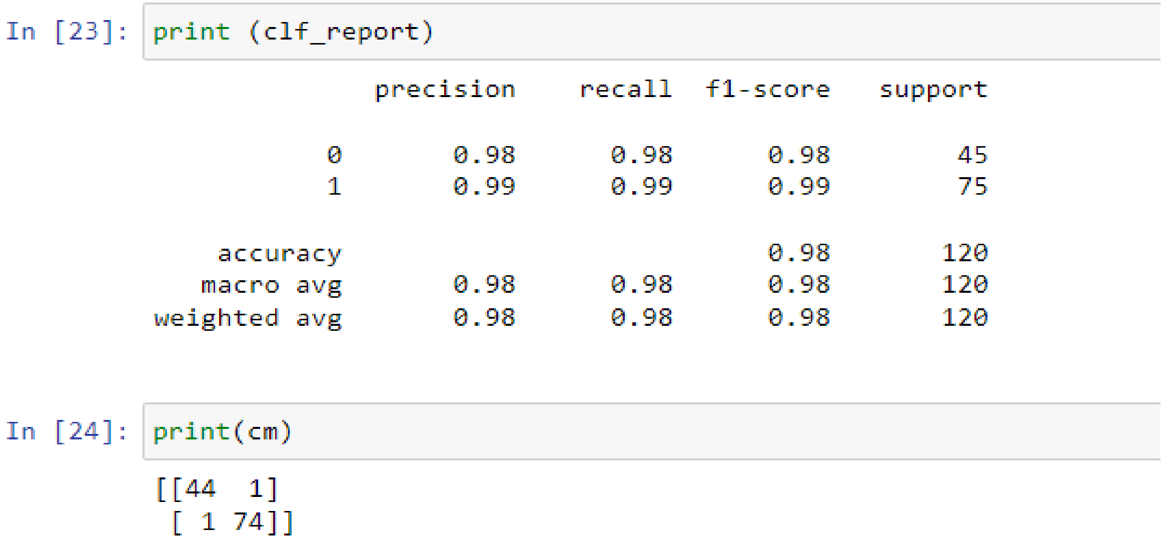
In this Dataset Label Encoder is used for preprocessing the data (converting text format to Numerical form)

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 finished all the algorithm and create all the models. I submit my all models through GITHUB

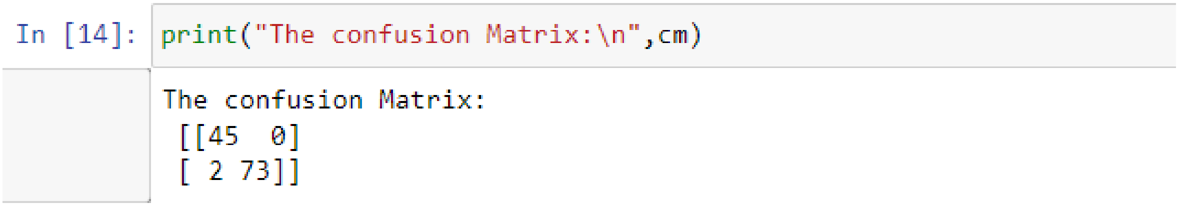
5.) All the research values of each algorithm should be documented. (You can make tabulation or screenshot of the results.)

Random Forest Classification:

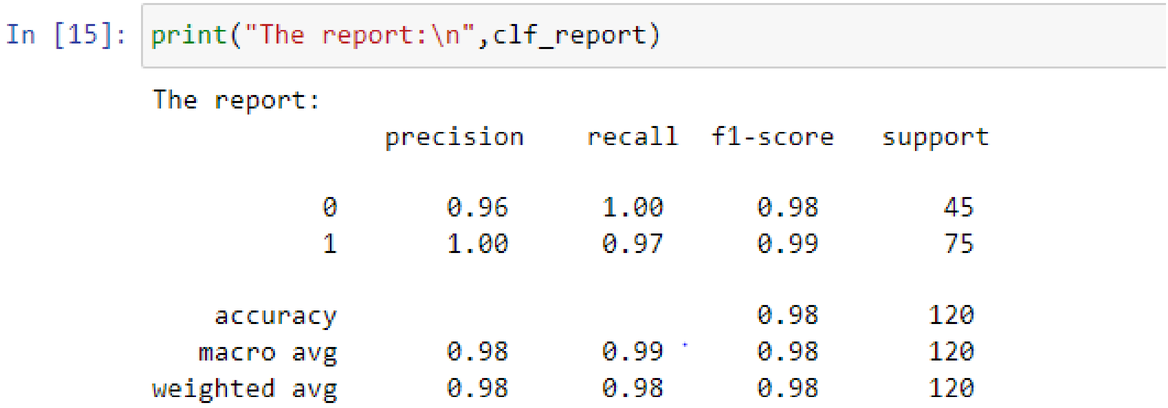


```
In [23]: print (clf_report)

                      precision    recall  f1-score   support

                 0         0.98      0.98      0.98        45
                 1         0.99      0.99      0.99        75

          accuracy                             0.98       120
         macro avg         0.98      0.98      0.98       120
      weighted avg         0.98      0.98      0.98       120

In [24]: print(cm)

         [[44  1]
          [ 1 74]]
```

Support Vector Machine Classification:

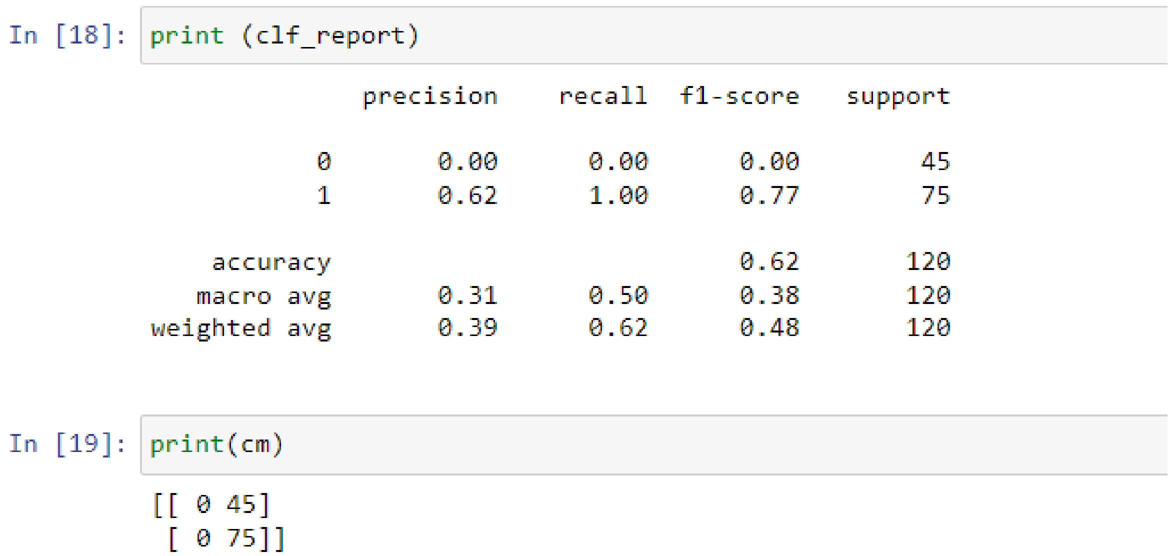


```
In [14]: print("The confusion Matrix:\n",cm)

         The confusion Matrix:
          [[45  0]
          [ 2 73]]
```



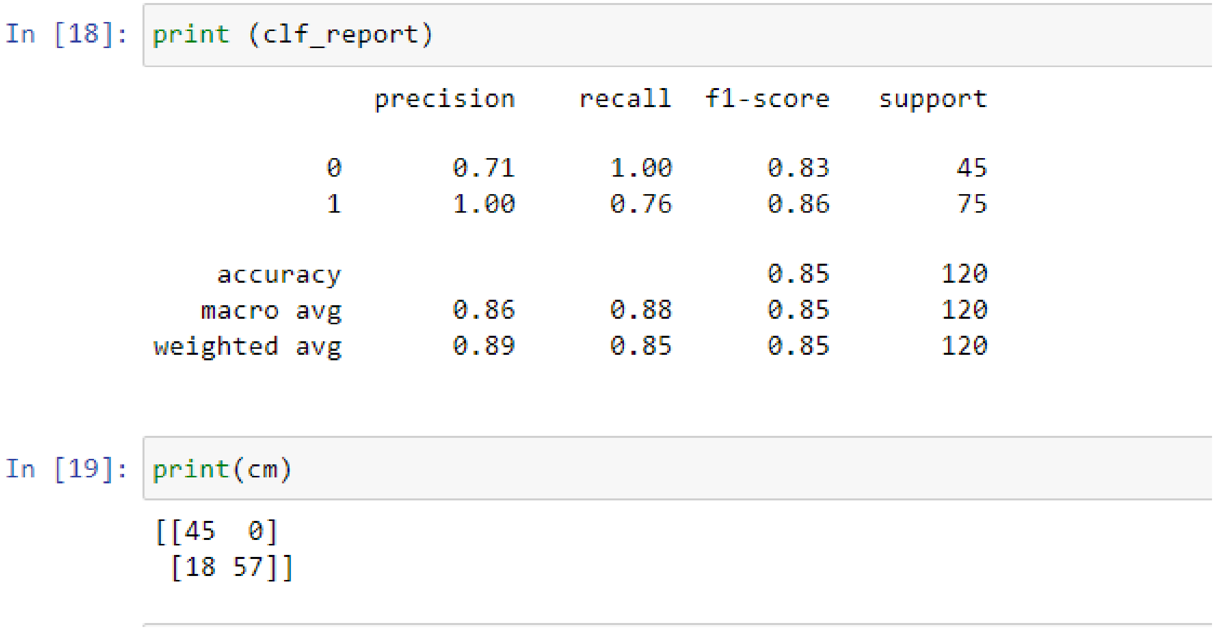
```
In [15]: print("The report:\n",clf_report)

         The report:
                       precision    recall  f1-score   support

                  0         0.96      1.00      0.98        45
                  1         1.00      0.97      0.99        75

           accuracy                             0.98       120
          macro avg         0.98      0.99      0.98       120
       weighted avg         0.98      0.98      0.98       120
```

Decision Tree Classification:



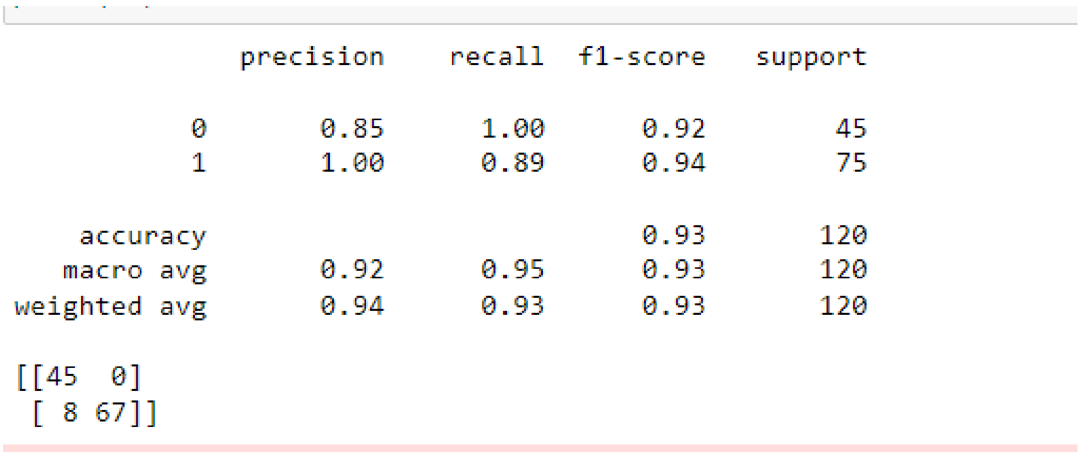
```
In [18]: print (clf_report)

              precision    recall  f1-score   support

           0       0.00      0.00      0.00        45
           1       0.62      1.00      0.77        75

    accuracy                           0.62       120
   macro avg       0.31      0.50      0.38       120
weighted avg       0.39      0.62      0.48       120

In [19]: print(cm)

[[ 0 45]
 [ 0 75]]
```

Logistics Regression Classification:



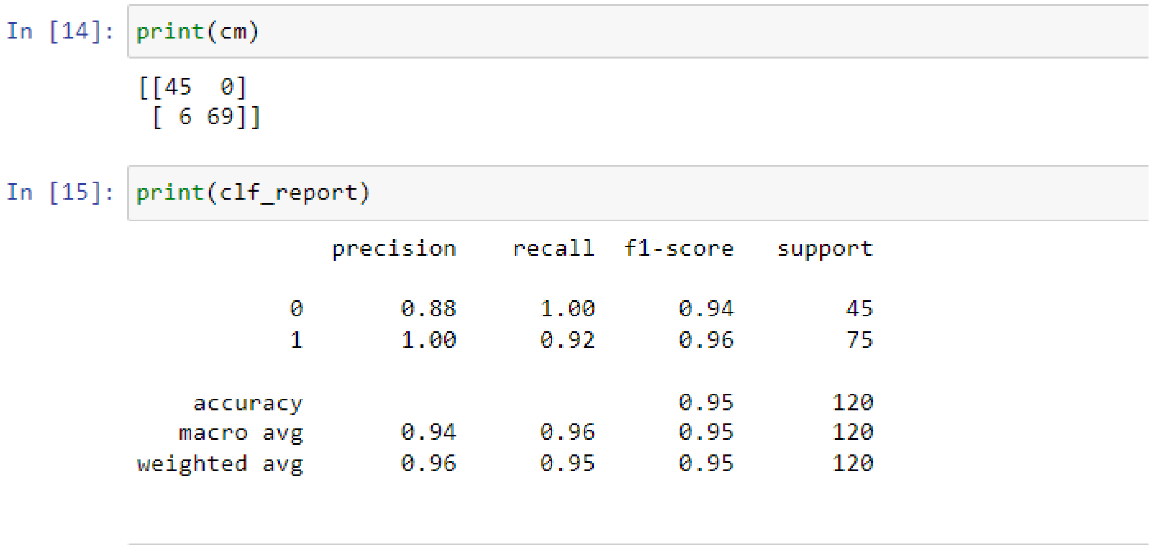
```
In [18]: print (clf_report)

              precision    recall  f1-score   support

           0       0.71      1.00      0.83        45
           1       1.00      0.76      0.86        75

    accuracy                           0.85       120
   macro avg       0.86      0.88      0.85       120
weighted avg       0.89      0.85      0.85       120

In [19]: print(cm)

[[45  0]
 [18 57]]
```

Naïve Bayes Classification:



```
              precision    recall  f1-score   support

           0       0.85      1.00      0.92        45
           1       1.00      0.89      0.94        75

    accuracy                           0.93       120
   macro avg       0.92      0.95      0.93       120
weighted avg       0.94      0.93      0.93       120

[[45  0]
 [ 8 67]]
```

KNN Classification:



```
In [14]: print(cm)

         [[45  0]
          [ 6 69]]

In [15]: print(clf_report)

                       precision    recall  f1-score   support

                    0       0.88      1.00      0.94        45
                    1       1.00      0.92      0.96        75

             accuracy                           0.95       120
            macro avg       0.94      0.96      0.95       120
         weighted avg       0.96      0.95      0.95       120
```

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

In this dataset the final model is the f1_macro value (Support Vector Machine Classification) for best parameter {'C': 10, 'gamma': 'auto', 'kernel': 'sigmoid'}: 0.98340188014101. Because other algo rithms f1_macro value is comparatively very low so I Choose the above model.