**CLASSIFICATION ASSIGNMENT**

**Problem Statement:**

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. Problem Identification:

* Stage – 1: Machine Learning
* Stage – 2: Supervised Learning
* Stage – 3: Classification –as the output label from the dataset are yes/no type

2. Basic info about the dataset:

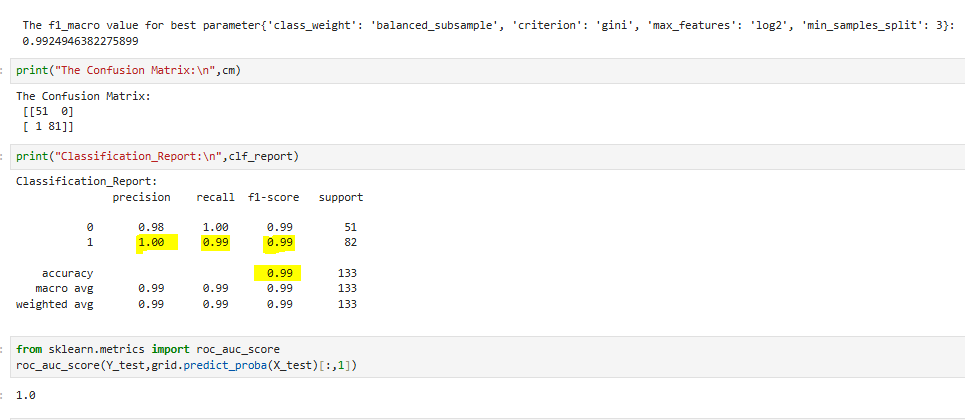
* Total number of rows : 399
* Total number of columns : 25

3. Preprocessing Method:

Preprocessing Technique like converting categorical data to nominal data by One- Hot Encoding for the columns **'rbc\_normal', 'pc\_normal','pcc\_present','ba\_present','htn\_yes','dm\_yes', 'cad\_yes','appet\_yes','pe\_yes','ane\_yes','classification\_yes'.**

4. Model Creation using Multiple Algorithms:

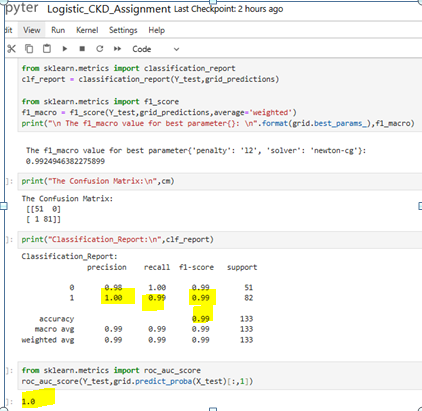
**i) Random Forest Classifier:**



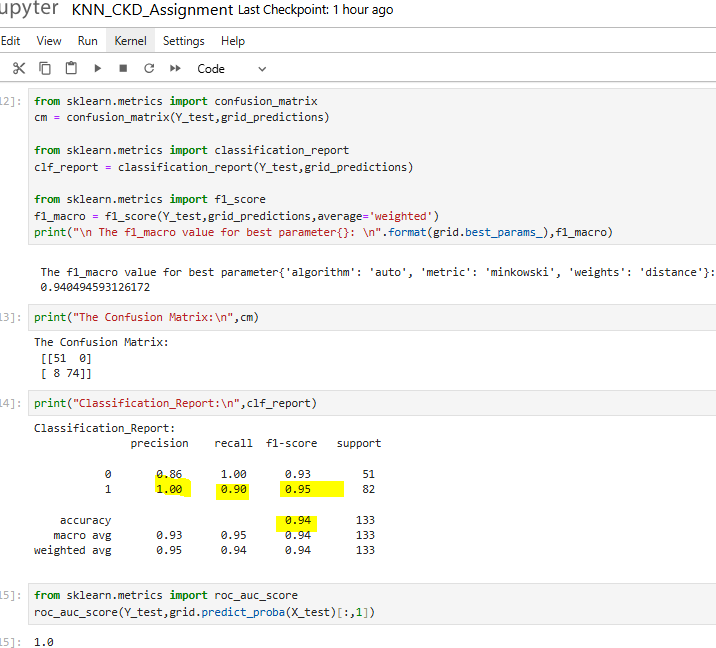
**ii) Decision Tree Classifier:**



**iii) Logistic Classifier:**



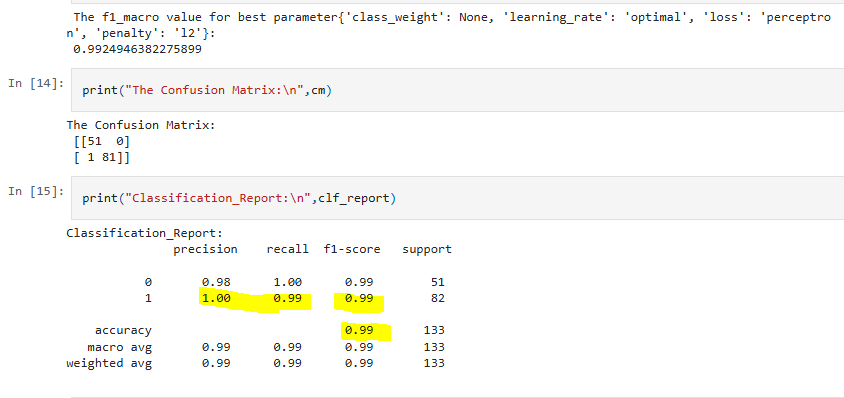
**iv) KNN algorithm:**



**V) SVM algorithm:**



**Vi) Stochastic Gradiant Descent Algorithm(SDG)**

****

**TABULATION BASED ON EVALUATION METRICS:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| S.No | Classifier Algorithm | Best Parameters | Precision | Recall | Accuracy | F1- score | ROC – AUC Score |
| 1 | Random Forest | 'class\_weight':'balanced\_subsample','criterion':'gini', 'max\_features':'log2', 'min\_samples\_split ': 3 | 1.00 | 0.99 | 0.99 | 0.99 | 1.00 |
| 2 | Decision Tree | 'criterion': 'log\_loss', 'max\_features': 'sqrt', 'splitter': 'random' | 0.99 | 0.95 | 0.96 | 0.97 | 0.96 |
| 3 | Support Vector Machine | 'decision\_function\_shape': 'ovo', 'gamma': 'scale', 'kernel': 'rbf' | 1.00 | 0.99 | 0.99 | 0.99 | 1.0 |
| 4 | K – Nearest Neighbor | 'algorithm': 'auto', 'metric': 'minkowski', 'weights': 'distance' | 1.00 | 0.90 | 0.94 | 0.95 | 1.0 |
| 5 | Logistic Regression | 'penalty': 'l2', 'solver': 'newton-cg' | 1.00 | 0.99 | 0.99 | 0.99 | 1.0 |
| 6 | Stochastic Gradiant Descent | 'class\_weight':None, 'learning\_rate':'optimal', 'loss': 'perceptron', 'penalty': 'l2' | 1.00 | 0.99 | 0.99 | 0.99 |  |

From the above analysis, the best method would be three algorithms gives the same results from Random Forest, Support Vector Machine, Logistic Regression, SGD.

For Reference please check my GitHub Link: <https://github.com/Geetharani-CodeAI/ML_Classification_Assignment-7>