**Classification Project**

**Requirement from Client:**

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. **Model Choosing:**

a. The input which is given by the client is in numbers, so we are going to use **Machine Learning**.

1. The Requirement from the Client is very clear, so in here we are going to use **Supervised Learning.**
2. The Label of the project is to predict the Chronic Kidney Disease(CKD), so we are going to use **Classification** for this project**.**

**2. Dataset Details:**

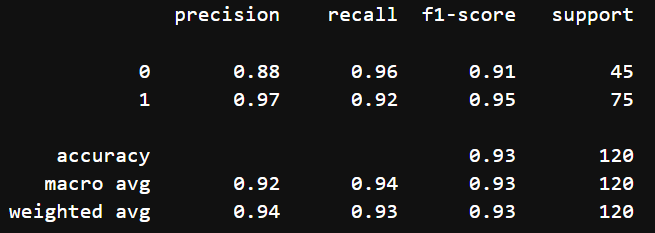
1. There are 400 Rows and 24 Columns in the Dataset

**3. Pre-Processing:**

a. In this dataset there are few string are available, so we are going to do preprocessing as nominal data.

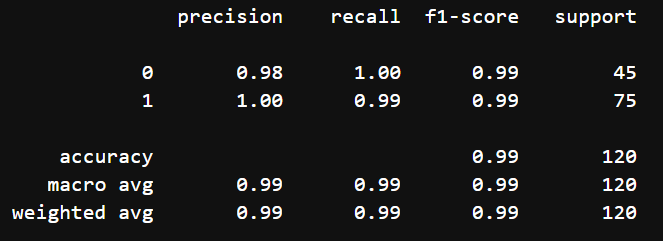
**4. Classification Algorithm’s Evaluation Metrics**

1. Decision Tree Classification:



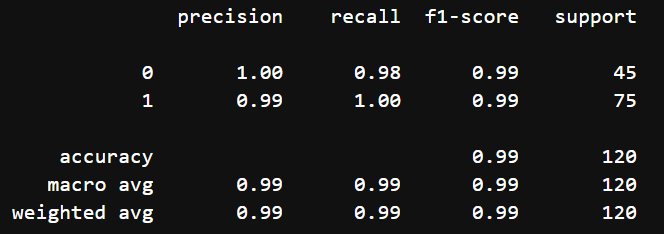
Criterion : entropy, max\_features : log2, splitter : best 🡪 is the best parameters in DT

1. Random Forest Classification:



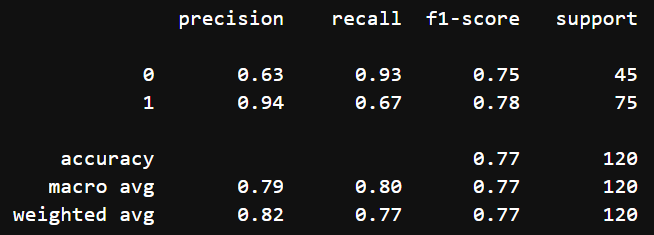
criterion: gini, max\_features: log2, n\_estimators: 50 🡪 is the best parameter in Random Forest.

1. Logistic Regression:



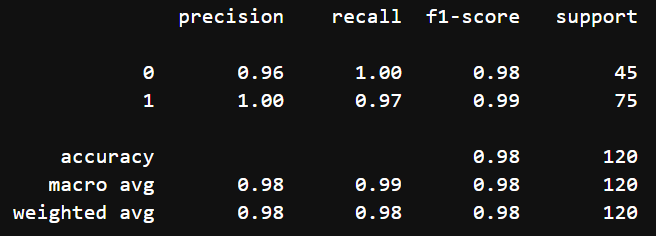
multi\_class : multinomial, penalty : l2, solver : newton-cg 🡪 is the best parameter in Logistic Regression

1. K Nearest Neighbors:

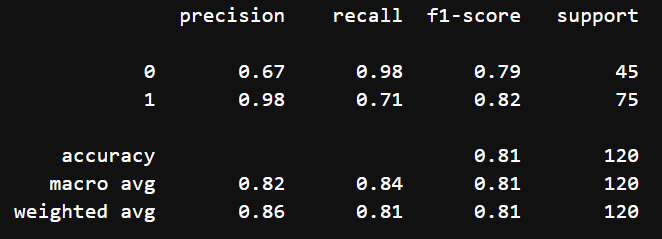


algorithm: auto, n\_neighbors: 5, weights: distance 🡪 is the best parameter in KNN

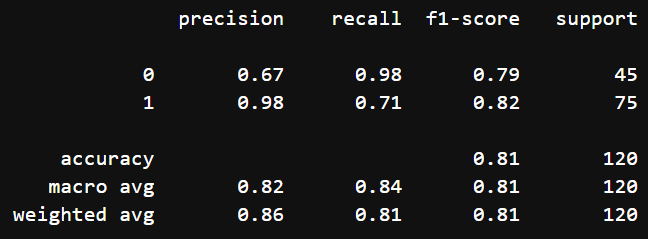
1. Naïve bayes:
2. GaussianNB



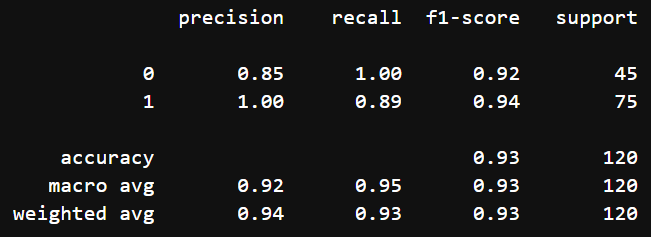
1. MultinomialNB



1. ComplementNB



1. BernoulliNB



**5. Final Model:**

1. By comparing all the Machine Learning Classification Algorithm’s Evaluation Metrics 🡪 Random Forest’s Accuracy has more Points(0.99), so RF is the best Algorithm,
2. (criterion: gini, max\_features: log2, n\_estimators: 50) 🡪 is the best parameter in RF.