**Classification Assignment**

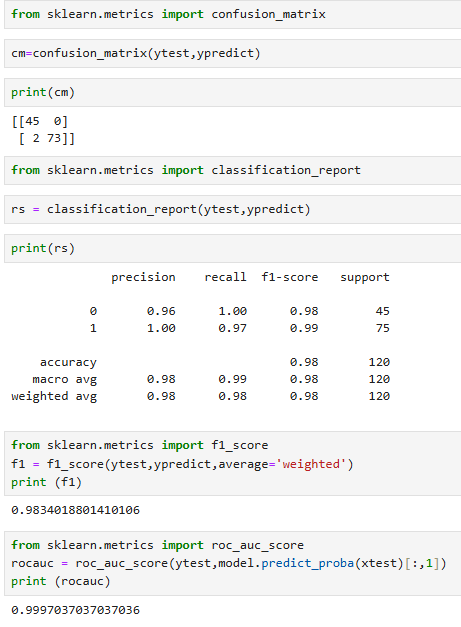
**Problem Statement:**

1. The given dataset contains 399 rows and 25 columns. In this the **classification column is our output** and the remaining columns are our input parameters.
2. The dataset contains alphabetical values that Machine language does not accept, so we have to convert the nominal data into numerical values by using **One Hot encoding**
3. We have a clear requirement, the client shared the proper dataset which contains multiple input values and one output as rows & column.

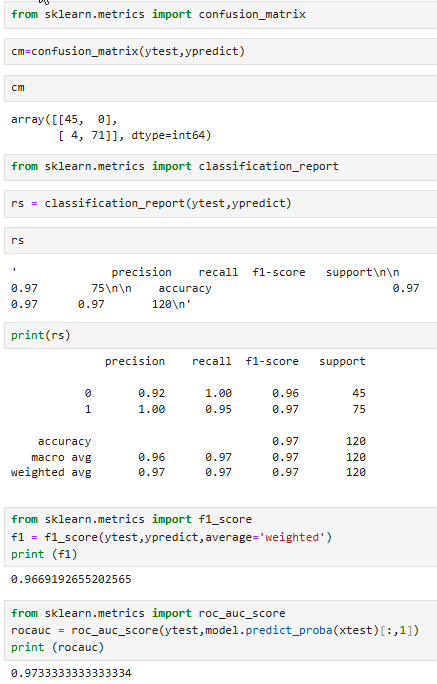
* The dataset has numerical values, hence we are dealing with numerical values this can come under **Machine Learning.**
* The dataset has proper input and output so it comes under **Supervised Learning**
* The end goal is to predict whether the patient will impact by Chronic Kidney Disease (CKD) or not, which is a classified output hence it comes under **Classification**

**Score:**

1. **SVM**



1. **Decision Tree**



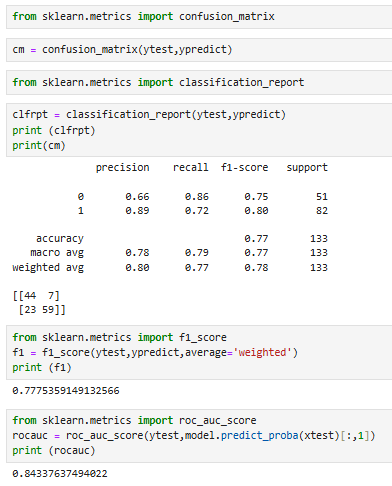
1. **Random Forest**



1. **Logistic Regression**

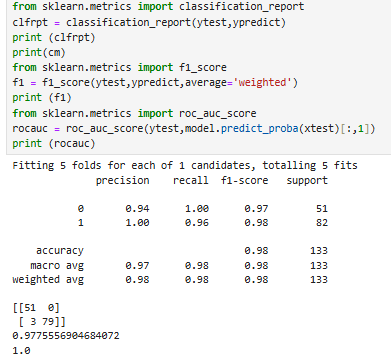


1. **KNN**

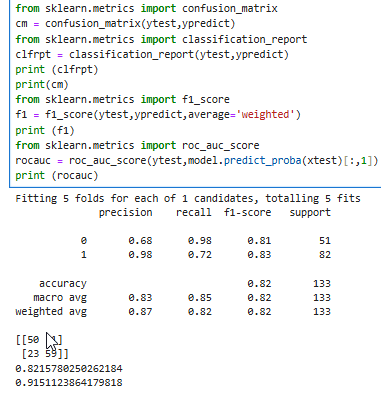


1. **Naive Bayes**

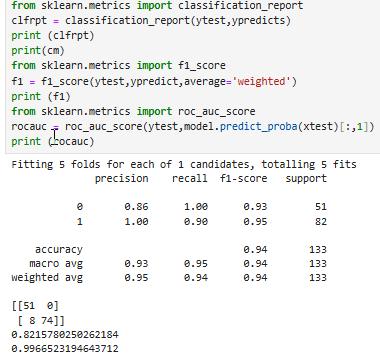
* **GaussianNB**



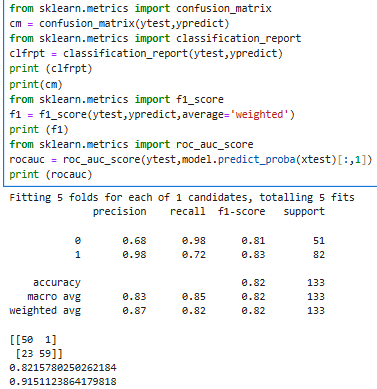
* **MultinomialNB**



* **BernoulliNB**



* **ComplementNB**



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sl No.** | **Algorithm** | **Accuracy** | **Recall** | **Precision** | **f1-score** | **roc\_auc** |
| 1 | SVM | 0.98 | 1.00 & 0.97 | 0.96 & 1.00 | 0.98 | 1.00 |
| 2 | Decision Tree | 0.97 | 1.00 & 0.95 | 0.92 & 1.00 | 0.96 | 0.97 |
| 3 | Random Forest | 0.98 | 0.98 & 0.99 | 0.98 & 0.99 | 0.98 | 1.00 |
| 4 | Logistic Regression | 0.98 | 0.96 & 1.00 | 1.00 & 0.98 | 0.98 | 1.00 |
| 5 | KNN | 0.77 | 0.86 & 0.72 | 0.66 & 0.89 | 0.77 | 0.84 |
| 6 | GaussianNB | 0.98 | 1.00 & 0.96 | 0.94 & 1.00 | 0.97 | 1.00 |
| 7 | MultinomialNB | 0.82 | 0.98 & 0.72 | 0.68 & 0.98 | 0.82 | 0.92 |
| 8 | BernoulliNB | 0.94 | 1.00 & 0.90 | 0.86 & 1.00 | 0.82 | 1.00 |
| 9 | ComplementNB | 0.82 | 0.98 & 0.72 | 0.68 & 0.98 | 0.82 | 0.92 |

1. Based on the above matrix report, considering f1-score we can choose SVM, Random Forest or Logistic Regression. But base on upward value in Recall I will choose SVM as a best model for this.