**Problem Statement:**

We need to create a model to predict the Chronic Kidney Disease (CKD) based on the several parameters provided by hospital.

**3 Stages of Problem Identification:**

**Stage 1:** Machine learning because we are dealing with numbers.

**Stage 2:** Supervised learning because we have both input and expected output.

**Stage 3:** Classification because we are classifying the input.

**Information about Dataset:**

**Total Rows:** 399

**Total Columns:** 28

**Yes Classification Rows:** 249

**No Classificstion Rows:** 150

**Input/Independent Variables:** age, bp, sg, al, su, rbc, pc, pcc, ba, bgr, bu,sc, sod, pot, hrmo, pcv, wc, rc, htn, dm, cad, appet, pe, ane

**Output/Dependent Variables:** Classification

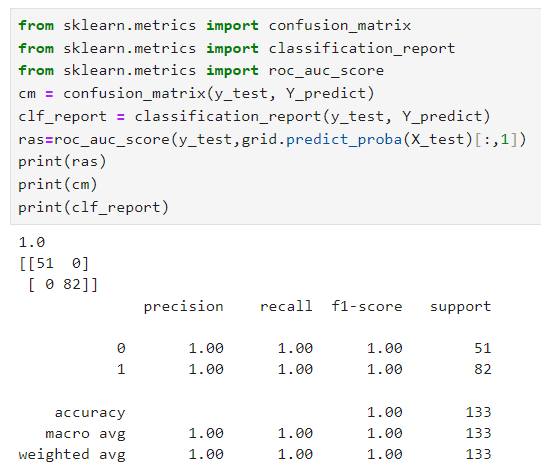
**Data Processing:**

'sg', 'rbc', 'pc', 'pcc', 'ba', 'htn', 'dm', 'cad', 'appet', 'pe', 'ane' , ‘classification’ are nominal data. It will be converted to numbers using getdummies(One hot encoding).

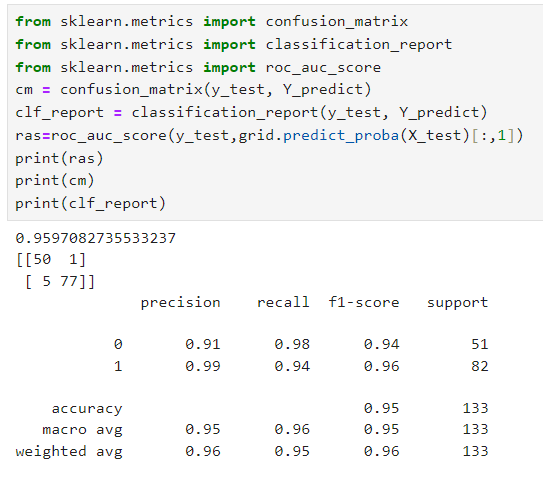
**Algorithm for AI Model:**

We can choose **Random Forest Regression Algorithm with parameters {'criterion': 'entrophy, 'max\_features': 'log2', 'n\_estimators': 50}** as a final model because it gave us the bestroc\_auc score, accuracy, Confusion matrix and classification report compare to other algorithms.

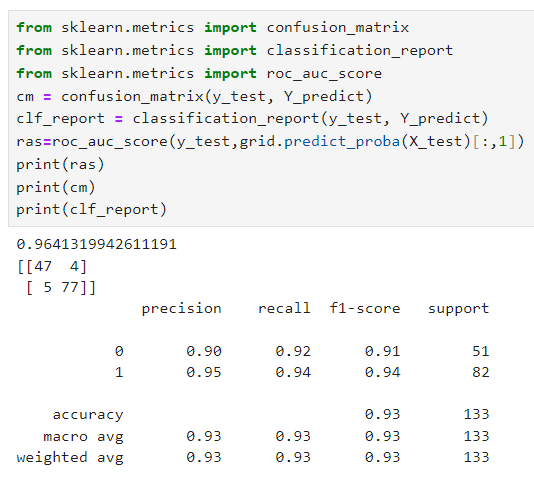
**Random Forest:**



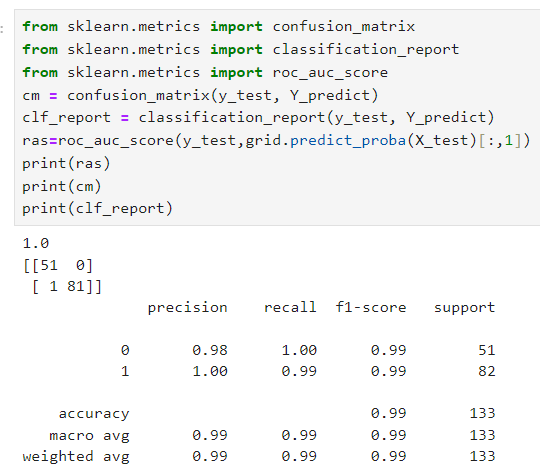
**DecisionTree:**



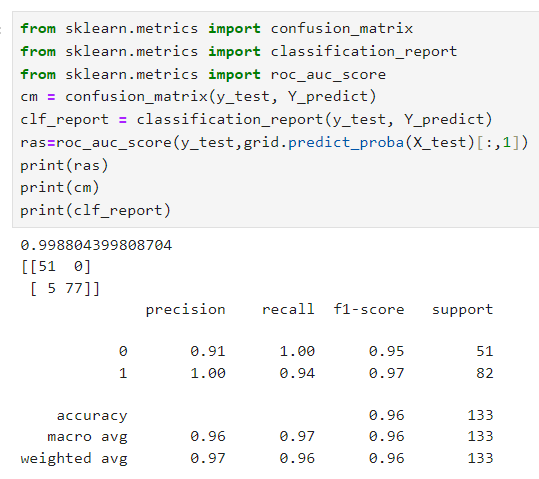
**SVM:**



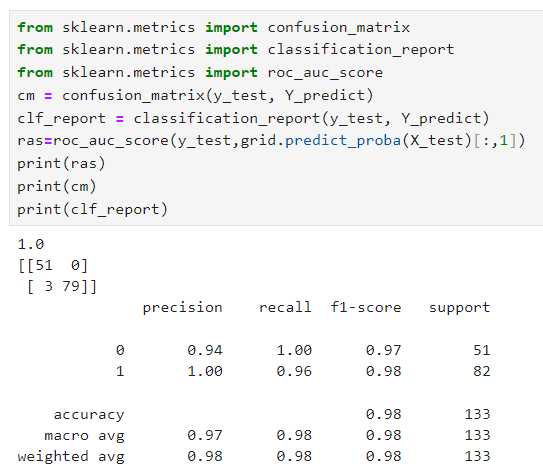
**LogisticRegression:**



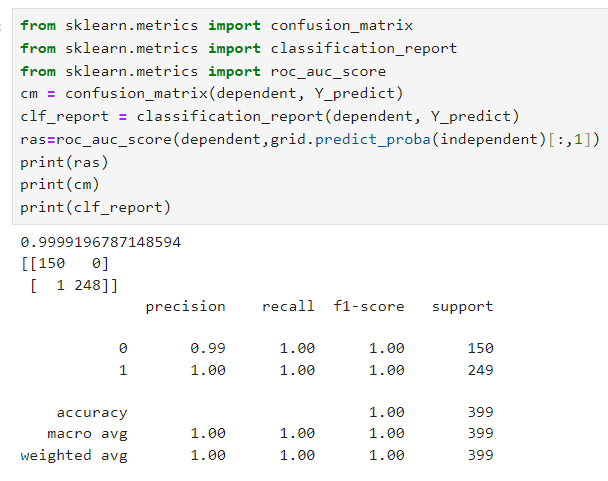
**KNN:**



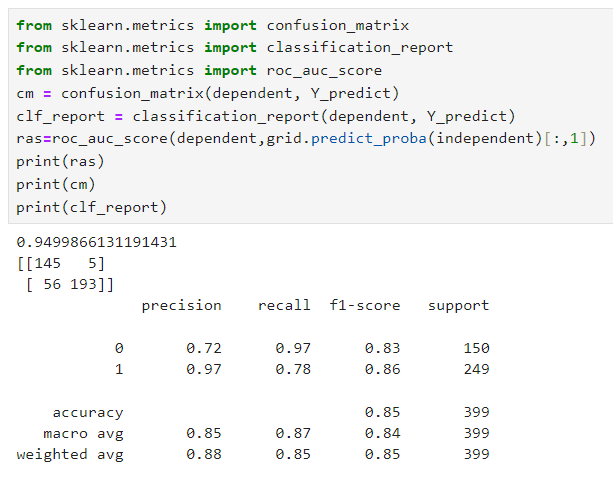
**BernoulliNaiveBayes:**



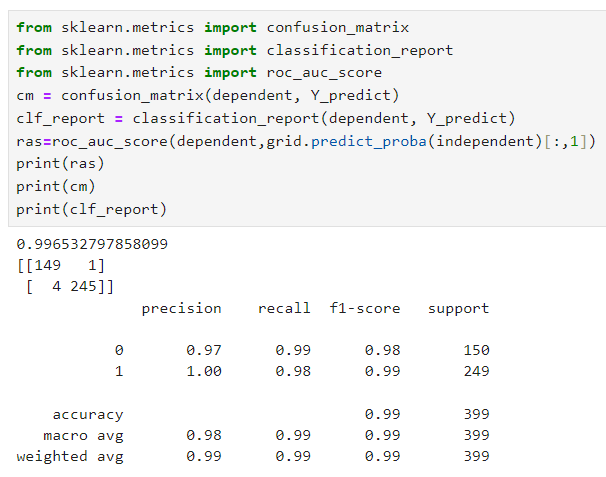
**CategoricalNaiveBayes:**



**ComplementNaiveBayes:**



**GaussianNaiveBayes:**



**MultinomialNaiveBayes:**

