**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

Based on a set of input parameters, we need to identify whether or not they will have chronic kidney disease.

* Input variables are numerical – so we use **Machine learning**
* We have clearly labelled data with a clear target variable of classification – so we use **Supervised learning**
* The required output is a binary outcome of Yes or no – so we use **Classification**

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

399 rows × 25 columns

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

Yes I converted categorical values to nominal values using:

dataset=pd.get\_dummies(dataset,dtype=int, drop\_first=True)

The pd.get\_dummies() function in pandas is used to perform one-hot encoding.

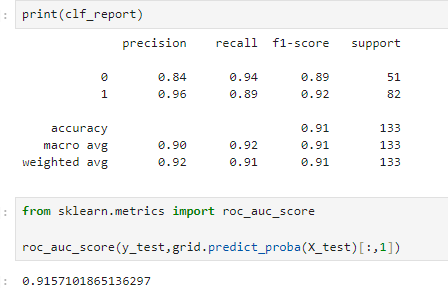
**One-Hot Encoding**: This is a technique used to convert categorical variables into a numerical format that can be used in machine learning algorithms. Each unique category is represented by a binary column (0 or 1).

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.

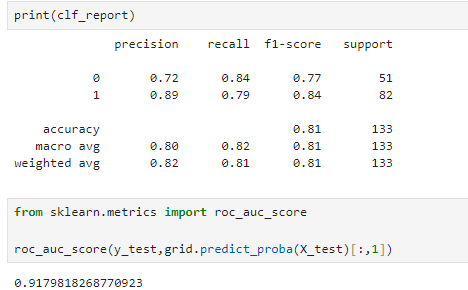
I ran 10 algorithms, out of which Random Forest was the best I have included the screenshots below.

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

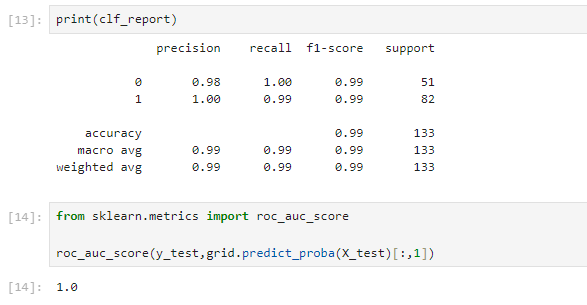
Decision Tree



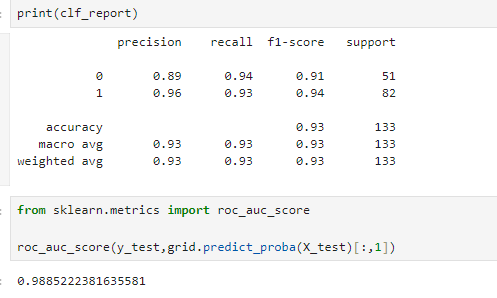
KNN



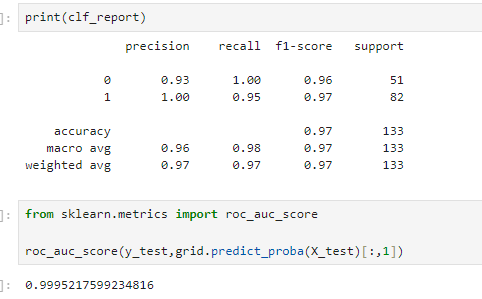
Random Forest



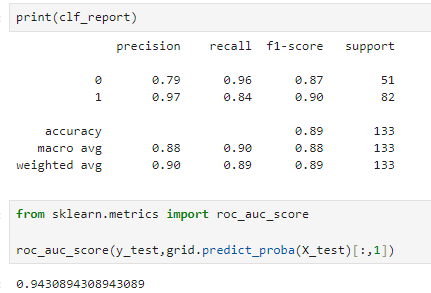
SVM



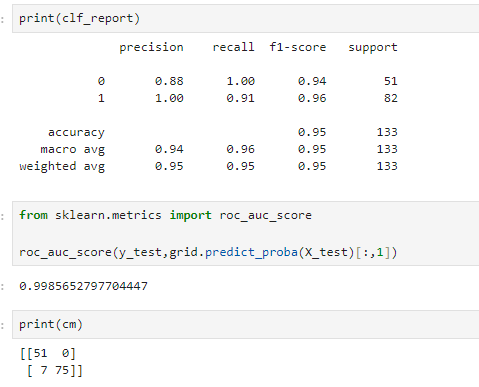
Logistic regression (with standard scaling of data)



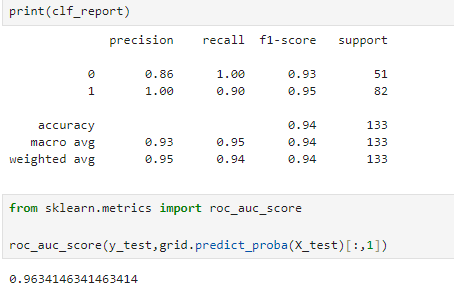
Multinomial NB



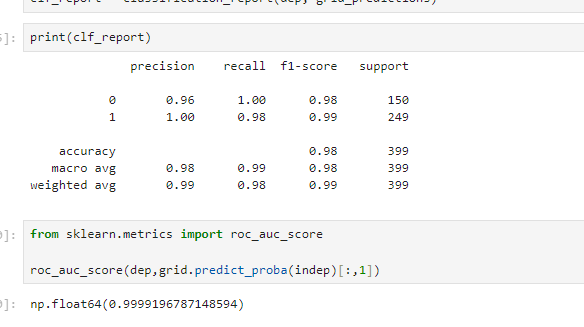
Gaussian NB



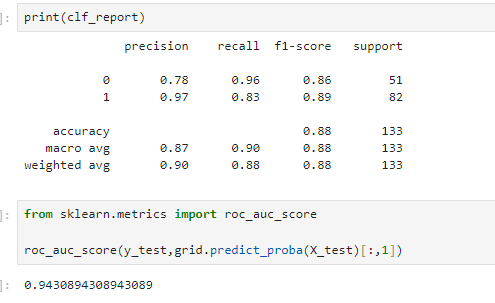
Bernoulli NB



Categorical NB



Complement NB



6.) Mention your final model, justify why u have chosen the same. Note: Mentioned points are necessary, kindly mail your document as well as .ipynb (code file) with respective name.

My final model pick is Random:

* It has an accuracy of 0.99 - This indicates that the model correctly classified 99% of the instances in the test dataset
* The roc\_auc value is a perfect 1 - This means that the ROC curve (Receiver Operating Characteristic curve) achieves the maximum value of 1, indicating that the model perfectly distinguishes between positive and negative classes across all thresholds

Also the best parameter value after tuning is as follows:

Best Parameters: {'bootstrap': False, 'criterion': 'entropy', 'max\_depth': None, 'min\_samples\_leaf': 1, 'min\_samples\_split': 2, 'n\_estimators': 100}