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

Predict the chronic kidney disease

Three Stages:

- Machine Learning
- Supervised Learning
- Classification

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

399 Rows and 25 Columns

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

Converted two rows of strings into nominal data, such as sg, rbc, pc, ba, htn, dm, cad, appet, pe, ane.

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.

Gradient Boosting Classifier is the good model.

```
The confusion matrix:
 [[51  0]
 [ 0 82]]
The report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        51
           1       1.00      1.00      1.00        82

    accuracy                           1.00       133
   macro avg       1.00      1.00      1.00       133
weighted avg       1.00      1.00      1.00       133

ROC AUC Score: 1.0
```

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

A) Decision Tree (ROC Score = 0.94)

```
print("The report:\n",clf_report)
```

```
The report:
               precision    recall  f1-score   support

           0       0.94      0.92      0.93        51
           1       0.95      0.96      0.96        82

    accuracy                           0.95       133
   macro avg       0.95      0.94      0.94       133
weighted avg       0.95      0.95      0.95       133
```

B) SVM (ROC Score = 1.00)

```
print("The report:\n",clf_report)
```

```
The report:
              precision    recall  f1-score   support

           0       0.96      1.00      0.98        51
           1       1.00      0.98      0.99        82

    accuracy                           0.98       133
   macro avg       0.98      0.99      0.98       133
weighted avg       0.99      0.98      0.99       133
```

C) Random Forest (ROC Score = 1.00)

```
print("The report:\n",clf_report)
```

```
The report:
              precision    recall  f1-score   support

           0       0.96      1.00      0.98        51
           1       1.00      0.98      0.99        82

    accuracy                           0.98       133
   macro avg       0.98      0.99      0.98       133
weighted avg       0.99      0.98      0.99       133
```

D) KNN (ROC Score = 0.98)

```
print("The report:\n",clf_report)
```

```
The report:
               precision    recall  f1-score   support

           0       0.91      1.00      0.95        51
           1       1.00      0.94      0.97        82

    accuracy                           0.96       133
   macro avg       0.96      0.97      0.96       133
weighted avg       0.97      0.96      0.96       133
```

E) Navie Bayes (Gaussian NB)

```
The confusion matrix:
 [[51  0]
 [ 3 79]]
The report:
              precision    recall  f1-score   support

           0       0.94      1.00      0.97        51
           1       1.00      0.96      0.98        82

    accuracy                           0.98       133
   macro avg       0.97      0.98      0.98       133
weighted avg       0.98      0.98      0.98       133

ROC AUC Score: 1.0
```

F) Navie Bayes (Multinomial NB)

```
The confusion matrix:
 [[50  1]
 [22 60]]
The report:
              precision    recall  f1-score   support

           0       0.69      0.98      0.81        51
           1       0.98      0.73      0.84        82

    accuracy                           0.83       133
   macro avg       0.84      0.86      0.83       133
weighted avg       0.87      0.83      0.83       133

ROC AUC Score: 0.9172644667623147
```

I) Navie Bayes (Bernoulli NB)

```
The confusion matrix:
 [[51  0]
 [ 9 73]]
The report:
              precision    recall  f1-score   support

           0       0.85      1.00      0.92        51
           1       1.00      0.89      0.94        82

    accuracy                           0.93       133
   macro avg       0.93      0.95      0.93       133
weighted avg       0.94      0.93      0.93       133

ROC AUC Score: 0.9991630798660928
```

J) Navie Bayes (Complement NB)

```
The confusion matrix:
 [[50  1]
 [22 60]]
The report:
              precision    recall  f1-score   support

           0       0.69      0.98      0.81        51
           1       0.98      0.73      0.84        82

    accuracy                           0.83       133
   macro avg       0.84      0.86      0.83       133
weighted avg       0.87      0.83      0.83       133

ROC AUC Score: 0.9172644667623147
```

K) Stochastic Gradient Descent

```
The confusion matrix:
 [[50  1]
 [ 1 81]]
The report:
              precision    recall  f1-score   support

           0       0.98      0.98      0.98        51
           1       0.99      0.99      0.99        82

    accuracy                           0.98       133
   macro avg       0.98      0.98      0.98       133
weighted avg       0.98      0.98      0.98       133

ROC AUC Score: 0.9995217599234816
```

L) Gaussian Processes

```
The confusion matrix:
 [[51  0]
 [ 2 80]]
The report:
              precision    recall  f1-score   support

           0       0.96      1.00      0.98        51
           1       1.00      0.98      0.99        82

    accuracy                           0.98       133
   macro avg       0.98      0.99      0.98       133
weighted avg       0.99      0.98      0.99       133

ROC AUC Score: 0.9997608799617408
```

M) Neural Network Model

```
The confusion matrix:
 [[51  0]
 [ 5 77]]
The report:
              precision    recall  f1-score   support

           0       0.91      1.00      0.95        51
           1       1.00      0.94      0.97        82

    accuracy                           0.96       133
   macro avg       0.96      0.97      0.96       133
weighted avg       0.97      0.96      0.96       133

ROC AUC Score: 0.9992826398852224
```

N) Extra Tree Classifier

```
The confusion matrix:
 [[51  0]
 [ 2 80]]
The report:
              precision    recall  f1-score   support

           0       0.96      1.00      0.98        51
           1       1.00      0.98      0.99        82

    accuracy                           0.98       133
   macro avg       0.98      0.99      0.98       133
weighted avg       0.99      0.98      0.99       133

ROC AUC Score: 1.0
```

O) Hist Gradient Boosting Classifier

```
The confusion matrix:
 [[51  0]
 [ 1 81]]
The report:
              precision    recall  f1-score   support

           0       0.98      1.00      0.99        51
           1       1.00      0.99      0.99        82

    accuracy                           0.99       133
   macro avg       0.99      0.99      0.99       133
weighted avg       0.99      0.99      0.99       133

ROC AUC Score: 0.9995217599234816
```

P) Gradient Boosting Classifier

```
The confusion matrix:
 [[51  0]
 [ 0 82]]
The report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        51
           1       1.00      1.00      1.00        82

    accuracy                           1.00       133
   macro avg       1.00      1.00      1.00       133
weighted avg       1.00      1.00      1.00       133

ROC AUC Score: 1.0
```

R) Bagging Classifier

```
The confusion matrix:
 [[50  1]
 [ 3 79]]
The report:
              precision    recall  f1-score   support

           0       0.94      0.98      0.96        51
           1       0.99      0.96      0.98        82

    accuracy                           0.97       133
   macro avg       0.97      0.97      0.97       133
weighted avg       0.97      0.97      0.97       133

ROC AUC Score: 0.9965327594452416
```

S) Voting Classifier

```
The confusion matrix:
 [[51  0]
 [ 2 80]]
The report:
              precision    recall  f1-score   support

           0       0.96      1.00      0.98        51
           1       1.00      0.98      0.99        82

    accuracy                           0.98       133
   macro avg       0.98      0.99      0.98       133
weighted avg       0.99      0.98      0.99       133

ROC AUC Score: 0.9995217599234816
The f1_macro value for best parameter {'lr__C': 100.0, 'rf__n_estimators': 20}: 0.9850141736106648
Prediction: [1]
```

T) Stacking Classifier

```
The confusion matrix:
 [[50  1]
 [ 1 81]]
The report:
              precision    recall  f1-score   support

           0       0.98      0.98      0.98        51
           1       0.99      0.99      0.99        82

    accuracy                           0.98       133
   macro avg       0.98      0.98      0.98       133
weighted avg       0.98      0.98      0.98       133

ROC AUC Score: 0.9997608799617408

array([1])
```

6. Mention your final model, justify why u have chosen the same.

Gradient Boosting Classifier is the good model. This model predicts accurately with an accuracy value of 1, and the ROC score is also 1.