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
2.) Tell basic info about the dataset (Total number of rows, columns)
3.) Mention the pre-processing method if you're doing any (like converting string to number – nominal data)
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.
5.) All the research values of each algorithm should be documented. (You can make tabulation or screenshot of the results.)
6.) Mention your final model, justify why u have chosen the same.

1.Machine learning---->Classification

2.Rows=399 and columns=28

3.Nominal data converting into True or False

4.create machine learning classification models are,

a.Logistic regression classification

b.Random Forest classification

c.Decision tree classification

d.KNN classification

e.Naive Bayes classification

5.All the research values

a.Logistic regression classification:

Roc_auc_score=1.0

```
the f1_macro value for best parameter{'penalty': 'l2', 'solver': 'newton-cg'}: 0.9916844900066377
the confusion matrix:
 [[45  0]
 [ 1 74]]
the report:
              precision    recall  f1-score   support

           0       0.98      1.00      0.99        45
           1       1.00      0.99      0.99        75

    accuracy                           0.99       120
   macro avg       0.99      0.99      0.99       120
weighted avg       0.99      0.99      0.99       120
```

b.Random Forest classification:

Roc_auc_score=0.99

```
the f1_macro value for best parameter{'criterion': 'entropy', 'max_features': 'log2', 'n_estimators': 100}: 0.9924946382275899
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
```

c.Decision tree classification:

Roc_auc_score=0.96

```
the f1_macro value for best parameter{'criterion': 'gini', 'max_features': 'sqrt', 'splitter': 'random'}: 0.9585802062760588
the confusion matrix:
 [[44  1]
 [ 4 71]]
the report:
              precision    recall  f1-score   support

       False       0.92      0.98      0.95        45
        True       0.99      0.95      0.97        75

    accuracy                           0.96       120
   macro avg       0.95      0.96      0.96       120
weighted avg       0.96      0.96      0.96       120
```

d.KNN classification:

Roc_auc_score=1.0

```
the f1_macro value for best parameter{'algorithm': 'auto', 'n_neighbors': 5, 'weights': 'distance'}: 0.9505208333333334
the confusion matrix:
 [[45  0]
 [ 6 69]]
the report:
              precision    recall  f1-score   support

       False       0.88      1.00      0.94        45
        True       1.00      0.92      0.96        75

    accuracy                           0.95       120
   macro avg       0.94      0.96      0.95       120
weighted avg       0.96      0.95      0.95       120
```

e.Naive Bayes classification:

Best model of naive bayes for GaussianNB and reason for best accuracy=0.98

```
              precision    recall  f1-score   support

       False       0.96      1.00      0.98        45
        True       1.00      0.97      0.99        75

    accuracy                           0.98       120
   macro avg       0.98      0.99      0.98       120
weighted avg       0.98      0.98      0.98       120

[[45  0]
 [ 2 73]]
```

6.The best final model Logistic regression classification model and reason for best accuracy=0.99 and roc_auc_score=1.0