PREDICTION OF CHRONIC KIDNEY DISEASE

1. **Problem Statement:**

   The Chronic Kidney Disease (CKD) dataset contains medical and clinical measurements used to determine whether a patient has CKD (Classification). The goal is to create a best model to predict the person/patient who has Chronic Kidney Disease based on features.

2. **Data Characteristics:**

   Target variable:

   Chronic Kidney Disease) - Yes (1)
   No Chronic Kidney Disease - No (0)

   Number of attributes:

   24 features - age, blood pressure, hemoglobin, serum creatinine, sodium, potassium, etc.

3. **Data Preprocessing:**

   Convert categorical data to numerical data using pd.get_dummies

4. **Final Model:**

   The Logistics Regression-classifier was selected due to its performance, high accuracy and roc-auc score.

Model 1 - Support Vector Classifier:

```
The Confusion Matrix:
 [[38  1]
 [ 0 41]]
```

```
The Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.97      0.99        39
           1       0.98      1.00      0.99        41

    accuracy                           0.99        80
   macro avg       0.99      0.99      0.99        80
weighted avg       0.99      0.99      0.99        80
```

The f1_macro value: 0.9874941323736504

Model 2 - Decision Tree Classifier:

```
The Confusion Matrix:
 [[38  1]
 [ 2 39]]
```

```
The Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.97      0.96        39
           1       0.97      0.95      0.96        41

    accuracy                           0.96        80
   macro avg       0.96      0.96      0.96        80
weighted avg       0.96      0.96      0.96        80
```

The f1_macro value: 0.9625058602906705

The ROC curve from prediction scores: 0.9627892432770482

Model 3 - Random Forest Classifier:

```
The Confusion Matrix:
 [[38  1]
 [ 2 39]]
```

```
The Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.97      0.99        39
           1       0.98      1.00      0.99        41

    accuracy                           0.99        80
   macro avg       0.99      0.99      0.99        80
weighted avg       0.99      0.99      0.99        80
```

The f1_macro value for best parameter {'criterion': 'gini', 'max_features': 'log2', 'n_estimators': 100}: 0.9749843652282676

The ROC curve from prediction scores: 1.0

Model 4 - Logistic Regression:

```
The confusion Matrix:
 [[32  0]
 [ 1 47]]
```

```
The Classification Report:
              precision    recall  f1-score   support

           0       0.97      1.00      0.98        32
           1       1.00      0.98      0.99        48

    accuracy                           0.99        80
   macro avg       0.98      0.99      0.99        80
weighted avg       0.99      0.99      0.99        80
```

The f1_macro value for best parameter {'C': 1.0, 'max_iter': 100, 'penalty': 'l2', 'solver': 'lbfgs'}: 0.9875303643724695

The ROC curve from prediction scores: 1.0