To Find the Best Classification Algorithm for the given problem statement

To Predict the Chronic Kidney Disease based on the given parameters from the Dataset

1) PROBLEM STAEMENT IDENTIFICATION:

STEP-1: Domain Selection- Machine Learning

STEP-2: Learning Selection- Supervised Learning

STEP-3: Classification

2) Basic information about the Dataset:

Total Number of Rows in the given Dataset=399

Total Number of Rows in the given Dataset=28

3) Pre-Processing Method:

In the given dataset columns consists of categorical value, hence converted to Nominal values (0 &1) by **One Hot Encoding** method using **get_dummies** function .

4) To find the best model in Machine Learning Classification Algorithm using Classification Report and ROC-AUC-Score value:

1.LOGISTIC REGRESSION: 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

       False       0.98      1.00      0.99        45
        True       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
```

2.SUPPORT VECTOR MACHINE Classification: ROC-AUC-SCORE=0.99

```
The f1-macro value for best parameter {'C': 10, 'gamma': 'auto', 'kernel': 'sigmoid'}: 0.9834018801410106
The confusion matrix:
 [[45  0]
 [ 2 73]]
The report:
              precision    recall  f1-score   support

           0       0.96      1.00      0.98        45
           1       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
```

3.**DECISION TREE Classification:** ROC-AUC-SCORE=0.95

```
The f1-macro value for best parameter {'criterion': 'gini', 'max_features': 'log2', 'splitter': 'random'}: 0.9503788982803847
The confusion matrix:
 [[44  1]
 [ 5 70]]
The report:
              precision    recall  f1-score   support

       False       0.90      0.98      0.94        45
        True       0.99      0.93      0.96        75

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

4.Random Forest Classification: ROC-AUC-SCORE=0.99

```
The f1-macro value for best parameter {'criterion': 'entropy', 'max_features': 'sqrt', 'n_estimators': 100}: 0.9916844900066377
The confusion matrix:
 [[45  0]
 [ 1 74]]
The report:
              precision    recall  f1-score   support

       False       0.98      1.00      0.99        45
        True       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
```

5.K-Nearest Neighbors (KNN) Classification: ROC-AUC-SCORE=1.0

```
The f1-macro value for best parameter {'algorithm': 'auto', 'metric': 'minkowski', 'n_neighbors': 3, 'p': 2}: 0.9587114337568058
The confusion matrix:
 [[45  0]
 [ 5 70]]
The report:
              precision    recall  f1-score   support

       False       0.90      1.00      0.95        45
        True       1.00      0.93      0.97        75

    accuracy                           0.96       120
   macro avg       0.95      0.97      0.96       120
weighted avg       0.96      0.96      0.96       120
```

Conclusion:

Hence for the given dataset for predicting Chronic Kidney Disease Random Forest Classification Algorithm predicts accurately than other algorithms of classification. Although other algorithms have higher accuracy and roc_auc_score but the prediction is not accurate, so RF Classification can be saved as best model.