Chronic Kidney Disease – ML Classification

PROBLEM STATEMENT:

Objective: Develop a machine learning model that accurately classifies patients as having **Chronic Kidney Disease (CKD)** or not, based on clinical and demographic data.

Context: CKD is a progressive condition that often remains undiagnosed until advanced stages, leading to high mortality and costly treatments. Early detection is critical to enable timely medical intervention and improve patient outcomes.

Data: The model will be trained on a dataset containing patient records with features such as serum creatinine, glomerular filtration rate (GFR), blood pressure, blood urea nitrogen (BUN), and other relevant indicators.

Goal: Improve diagnostic accuracy and support clinical decision-making by identifying key predictors of CKD. Performance will be evaluated using metrics such as accuracy, precision, recall, and F1-score.

Impact: A reliable classification model in Machine Learning can assist healthcare providers in early screening, reduce diagnostic delays, and potentially lower the burden of CKD through proactive treatment strategies.

TOTAL NUMBER OF ROWS AND COLUMNS: 399 rows * 29 columns

TYPES OF CLASSIFICATION USED:

1. **RANDOM FOREST GRID CLASSIFICATION**
2. **SVM GRID CLASSIFICATION**
3. **DECISION TREE GRID CLASSIFICATION**
4. **LOGISTIC GRID CLASSIFICATION**

CLASSIFICATION MODEL	Fit Model	F1_Score	Confusion Matrix	Roc_auc_score
1. **RANDOM FOREST GRID CLASSIFICATION**	Fitting 5 folds for each of 12 candidates, totalling 60 fits	0.9109616145627226	[[51 0] [12 70]]	-
2. **SVM GRID CLASSIFICATION**	Fitting 5 folds for each of 40 candidates, totalling 200 fits	0.8961766485178463	The confusion Matrix: [[51 0] [14 68]]	0.9024390243902438
3. **DECISION TREE GRID CLASSIFICATION**	Fitting 5 folds for each of 12 candidates, totalling 60 fits	0.9035727816517972	The confusion Matrix: [[51 0] [13 69]]	0.9268292682926829
4. **LOGISTIC GRID CLASSIFICATION**	Fitting 5 folds for each of 4 candidates, totalling 20 fits	0.9109616145627226	The confusion Matrix: [[51 0] [12 70]]	0.9268292682926829

Among all the models the SVM GRID CLASSIFICATION is the best model for this CKD prediction.

1.RANDOM FOREST GRID CLASSIFICATION

RF Grid: 399 rows × 28 columns

Fit Model: Fitting 5 folds for each of 12 candidates, totalling 60 fits

```
Fitting 5 folds for each of 12 candidates, totalling 60 fits
```

f1_macro: 0.9109616145627226

```
The f1_macro value for best parameter {'criterion': 'entropy', 'max_features': 'sqrt', 'n_estimators': 10}: 0.9109616145627226
```

Confusion Matrix:

```
The confusion Matrix:
 [[51  0]
 [12 70]]
```

Classification Report:

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

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

           0       0.81      1.00      0.89        51
           1       1.00      0.85      0.92        82

    accuracy                           0.91       133
   macro avg       0.90      0.93      0.91       133
weighted avg       0.93      0.91      0.91       133
```

2. SVM GRID CLASSIFICATION

Total Number of Rows and Columns: 399 rows × 28 columns

Fit Model:

```
Fitting 5 folds for each of 40 candidates, totalling 200 fits
```

F1_Macro:

```
The f1_macro value for best parameter {'C': 2000, 'gamma': 'auto', 'kernel': 'sigmoid'}: 0.8961766485178463
```

Confusion Matrix:

```
The confusion Matrix:
 [[51  0]
 [14 68]]
```

Classification Report:

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

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

           0       0.78      1.00      0.88        51
           1       1.00      0.83      0.91        82

    accuracy                           0.89       133
   macro avg       0.89      0.91      0.89       133
weighted avg       0.92      0.89      0.90       133
```

Roc_auc_score:

```
roc_auc_score(y_test, grid.decision_function(X_test))
```

```
0.9024390243902438
```

3.DECISION TREE GRID CLASSIFICATION

Total Number of Rows and Columns: 399 rows × 28 columns

Fit Model:

```
Fitting 5 folds for each of 12 candidates, totalling 60 fits
```

F1_macro:

```
The f1_macro value for best parameter {'criterion': 'gini', 'max_features': 'sqrt', 'splitter': 'random'}: 0.9035727816517972
```

Confusion Matrix:

```
The confusion Matrix:
 [[51  0]
 [13 69]]
```

Classification Report:

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

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

           0       0.80      1.00      0.89        51
           1       1.00      0.84      0.91        82

    accuracy                           0.90       133
   macro avg       0.90      0.92      0.90       133
weighted avg       0.92      0.90      0.90       133
```

Roc_auc_score:

```
roc_auc_score(y_test,grid.predict_proba(X_test)[:,1])
```

```
0.9268292682926829
```

4.LOGISTIC GRID CLASSIFICATION

Total Number of Rows and Columns: 399 rows* 28 Columns

Fit Model:

```
Fitting 5 folds for each of 4 candidates, totalling 20 fits
```

F1_macro value:

```
The f1_macro value for best parameter {'penalty': 'l2', 'solver': 'newton-cg'}: 0.9109616145627226
```

Confusion Matrix:

```
The confusion Matrix:
 [[51  0]
 [12 70]]
```

Classification Report:

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

           0       0.81      1.00      0.89        51
           1       1.00      0.85      0.92        82

    accuracy                           0.91       133
   macro avg       0.90      0.93      0.91       133
weighted avg       0.93      0.91      0.91       133
```

Roc_auc_score:

```
roc_auc_score(y_test,grid.predict_proba(X_test)[:,1])
```

```
0.9268292682926829
```