

Classification Assignment

1. Problem Statement:

Develop a predictive machine learning model that identifies Chronic Kidney Disease (CKD) in patients using clinical parameters provided by hospital management. The goal is to assist medical staff in early diagnosis and effective management of CKD.

2. Dataset Overview:

- 399 rows × 28 columns

3. Encoding Categorical Data:

Converted columns from string to numbers using one-hot encoding.

Handling Missing Data:

Verified there are no missing values; ready for modeling.

4. Model Development and Evaluation:

Experimented with the following machine learning algorithms using Python's sklearn and libraries.

5. The research values:

- Logistic Regression - GRID

```
[17]: print("The confusion Matrix:\n",cm)
The confusion Matrix:
[[51  0]
 [ 1 81]]

[18]: print("The report:\n",clf_report)
The report:
      precision    recall  f1-score   support

   False      0.98      1.00      0.99         51
    True      1.00      0.99      0.99         82

 accuracy          0.99
 macro avg          0.99
weighted avg          0.99
```

- Decision Tree-GRID:

```
[16]: print("The confusion Matrix:\n",cm)
```

```
The confusion Matrix:
[[50  1]
 [ 1 81]]
```

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

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

   False      0.98      0.98      0.98        51
   True       0.99      0.99      0.99        82

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

- Random Forest-GRID:

```
print("The confusion Matrix:\n",cm)
```

```
The confusion Matrix:
[[ 0 51]
 [ 0 82]]
```

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

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

   False      0.00      0.00      0.00        51
   True       0.62      1.00      0.76        82

 accuracy      0.62      0.62      0.62       133
 macro avg     0.31      0.50      0.38       133
 weighted avg  0.38      0.62      0.47       133
```

- Support Vector Machine (SVM)-GRID

```
print("The confusion Matrix:\n",cm)
```

```
The confusion Matrix:
[[51  0]
 [ 2 80]]
```

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

6. Final Model Selection:

Logistic Regression -GRID – Classifier has the highest overall accuracy (99%) and excellent precision, recall, and F1 scores balanced perfectly across classes.