Classification Assignment

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
- **Machine Learning – related to Numbers**
- **Supervised Learning – Both input and output are known**
- **Classification Method – Categorical data to Numerical data**

2.) Tell basic info about the dataset (Total number of rows, columns)
- **399 rows**
- **25 columns**

3.) Mention the pre-processing method if you're doing any (like converting string to number – nominal data)
- **One-hot Encoding**

4.) Used evaluation metric
- confusion_Matrix
- f1_score
- Roc_auc_score
- classification_report

5.) Used machine learning algorithm for classification Assignment
- RF –Random Forest
- DT– Decision Tree
- SVM – Support Vector Machine
- LOG – Logistic Regression
- KNN - K-Nearest Neighbors
- NaviesBayes (GaussianNB, MultinomialNB, BernoulliNB, ComplementNB)

Random Forest results:

	precision	recall	f1-score	support
False	0.98	0.98	0.98	51
True	0.99	0.99	0.99	82
accuracy			0.98	133
macro avg	0.98	0.98	0.98	133
weighted avg	0.98	0.98	0.98	133

Decision Tress results:

	precision	recall	f1-score	support
False	0.94	1.00	0.97	51
True	1.00	0.96	0.98	82
accuracy			0.98	133
macro avg	0.97	0.98	0.98	133
weighted avg	0.98	0.98	0.98	133

Support Vector Machine results:

	precision	recall	f1-score	support
False	0.98	0.98	0.98	51
True	0.99	0.99	0.99	82
accuracy			0.98	133
macro avg	0.98	0.98	0.98	133
weighted avg	0.98	0.98	0.98	133

Logistic Regression results:

	precision	recall	f1-score	support
False	0.89	0.92	0.90	51
True	0.95	0.93	0.94	82
accuracy			0.92	133
macro avg	0.92	0.92	0.92	133
weighted avg	0.93	0.92	0.93	133

K-Nearest Neighbors results:

GaussianNB

	precision	recall	f1-score	support
False	0.94	1.00	0.97	51
True	1.00	0.96	0.98	82
accuracy			0.98	133
macro avg	0.97	0.98	0.98	133
weighted avg	0.98	0.98	0.98	133

MultinomialNB

	precision	recall	f1-score	support
False	0.68	0.98	0.81	51
True	0.98	0.72	0.83	82
accuracy			0.82	133
macro avg	0.83	0.85	0.82	133
weighted avg	0.87	0.82	0.82	133

BernoulliNB

	precision	recall	f1-score	support
False	0.86	1.00	0.93	51
True	1.00	0.90	0.95	82
accuracy			0.94	133
macro avg	0.93	0.95	0.94	133
weighted avg	0.95	0.94	0.94	133

ComplementNB

	precision	recall	f1-score	support
False	0.68	0.98	0.81	51
True	0.98	0.72	0.83	82
accuracy			0.82	133
macro avg	0.83	0.85	0.82	133
weighted avg	0.87	0.82	0.82	133

6.)Used Hypertunning parameter for below Algorithm

- RF –Random Forest
- DT– Decision Tree
- SVM – Support Vector Machine
- LOG – Logistic Regression

RF-Random Forest

F1_score: 0.99

The classification 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	133
macro avg	0.99	0.99	0.99	133
weighted avg	0.99	0.99	0.99	133

Roc_auc_score: 0.99

DT– Decision Tree

F1_score: 0.93

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

       False       0.86      0.98      0.92        51
        True       0.99      0.90      0.94        82

    accuracy                           0.93       133
   macro avg       0.92      0.94      0.93       133
weighted avg       0.94      0.93      0.93       133
```

Roc_auc_score: 0.94

SVM – Support Vector Machine

F1_score: 0.99

```
The classification 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       133
   macro avg       0.99      0.99      0.99       133
weighted avg       0.99      0.99      0.99       133
```

Roc_auc_score: 1

LOG – Logistic Regression

F1_score:0.99

```
The classification 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       133
   macro avg       0.99      0.99      0.99       133
weighted avg       0.99      0.99      0.99       133
```

Roc_auc_score: 1

Final Best Model: Grid-Logistic Regression

Both the tuned Grid SVM and Grid Logistic Regression models produced nearly identical performance. Since there was no meaningful difference in their results, I selected the tuned Logistic Regression model. We evaluated multiple prediction models, and these two performed equally well. Logistic Regression was chosen because it is simpler to understand, easier to explain, and more straightforward to deploy in real-world applications.