Classification Assignment:

1. Identify your problem statement.

   Customer wanted to predate the Chronic Kidney Disease(CKD) based on several inputs.

   Stage1: Machin Learning

   Stage2: Supervised Learning

   Stage3: Classifications

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

Total 399 rows × 28 columns, in that last column i.e: classification need to predicte

3. Develop a good model with good evaluation metric. You can use any machine learning algorithm; you can create many models. Finally, you have to come up with final model.

Results from all other algorithms:

Algorithm types	Type I Error	Type 2 Error	Accuracy	Best Parameter	Roc_auc_score
Random Forest	0	0	1.0	{'criterion': 'log_loss', 'max_features': 'log2', 'n_estimators': 10}	1.0
Decision Tree	1	3	0.97	{'criterion': 'log_loss', 'max_features': 'sqrt', 'splitter': 'random'}	0.971903395
KNN	7	25	0.76	{'algorithm': 'auto', 'n_neighbors': 8, 'weights': 'distance'}	0.811095169
Logistics	1	0	0.99	{'C': 1.0, 'multi_class': 'multinomial', 'penalty': 'l2', 'solver': 'newton-cg'}	1.0
Navie Bayes	0	8	0.94	Default	0.996652

BernoulliNB					
Navie Bayes ComplementNB	1	23	0.82	Default	0.915112386
Navie Bayes MultinomialNB	1	23	0.82	Default	0.9151123
Navie Bayes GaussianNB	0	3	0.98	Default	1.0

Conclusion:

Based on algorithm analysis I am getting best score on Random Forest and Logistics algorithm,

In Random Forest: we don't have an Error with Type1 & 2 , it gives the best score.

In Navie BayeS GaussianNB: Type1 error is minimum as compared with Type2 error

I would go with **Navie BayeS GaussianNB is best model for this problem.**

naive_bayes: BernoulliNB

array([[51, 0],

[8, 74]], dtype=int64)

roc_auc_score: 0.9966523194643712

CategoricalNB:

array([[51, 0],

[8, 74]], dtype=int64)

roc_auc_score: 0.9966523194643712

GaussianNB:

array([[51, 0],

[3, 79]], dtype=int64)

roc_auc_score:

1.0

MultinomialNB:

```
array([[50,  1],
       [23, 59]], dtype=int64)
```

roc_auc_score:

0.9151123864179818