# Classification Assignment

## Problem Statement Analysis

The Chronic Kidney Disease (CKD) dataset includes several features, with the "classification" field being the target variable to predict. The other fields, totaling 24 input features, are used as predictors to train and test the model.

## Preprocessing Method

The dataset contains categorical fields that must be converted into numerical values for effective model processing. This transformation is achieved using a label encoder, which converts categorical data into integers.

## Model Results

### Random Forest Results:

```
Fitting 5 folds for each of 12 candidates, totalling 60 fits
Best Parameters: {'criterion': 'entropy', 'n_estimators': 1000}
              precision    recall  f1-score   support

           0       0.98      0.98      0.98        51
           1       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 Tree:

```
Fitting 5 folds for each of 6 candidates, totalling 30 fits
Best Parameters: {'criterion': 'gini', 'splitter': 'random'}
              precision    recall  f1-score   support

           0       0.86      0.98      0.92        51
           1       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
```

## Support Vector Machine

```
Fitting 5 folds for each of 3 candidates, totalling 15 fits
Best Parameters: {'kernel': 'sigmoid', 'random_state': 0}
              precision    recall  f1-score   support

           0       0.27      0.12      0.16        51
           1       0.59      0.80      0.68        82

    accuracy                           0.54       133
   macro avg       0.43      0.46      0.42       133
weighted avg       0.47      0.54      0.48       133
```

## Logistic Regression for classification

```
Fitting 5 folds for each of 9 candidates, totalling 45 fits
Best Parameters: {'C': 10, 'penalty': 'l2', 'solver': 'liblinear'}
              precision    recall  f1-score   support

           0       0.89      0.92      0.90        51
           1       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
```

## KNN

```
Fitting 5 folds for each of 60 candidates, totalling 300 fits
Best Parameters: {'metric': 'cosine', 'n_neighbors': 5, 'p': 1}
              precision    recall  f1-score   support

           0       0.80      0.94      0.86        51
           1       0.96      0.85      0.90        82

    accuracy                           0.89       133
   macro avg       0.88      0.90      0.88       133
weighted avg       0.90      0.89      0.89       133
```

## Navie Baiyas: MultinomialNB

```
Fitting 5 folds for each of 10 candidates, totalling 50 fits
Best Parameters: {'alpha': 0.1, 'fit_prior': True}
              precision    recall  f1-score   support

           0       0.98      1.00      0.99        51
           1       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
```

**BernoulliNB**

```
Fitting 5 folds for each of 10 candidates, totalling 50 fits
Best Parameters: {'alpha': 0.5, 'fit_prior': False}
              precision    recall  f1-score   support

           0       0.96      1.00      0.98        51
           1       1.00      0.98      0.99        82

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

**CategoricalNB**

```
Best Parameters: {'alpha': 0.1, 'fit_prior': True}
              precision    recall  f1-score   support

           0       0.98      1.00      0.99        51
           1       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
```

## Final Model

Based on the analysis, both **Naive Bayes MultinomialNB** and **CategoricalNB** demonstrate superior performance scores.