# Classification Algorithm Report

## 1. Objective

Predict Chronic Kidney Disease (CKD) using AI based on the following features: age, bp, al, su, bgr, bu, sc, sod, pot, hrmo, pcv, wc, etc.

## 2. Dataset

The dataset has dimensions of 399 × 25 (399 rows and 25 columns). Both input features and the target (classification column) are present. The dataset is quite imbalanced.

## 3. Preprocessing

Since the dataset includes categorical variables, one-hot encoding was applied.

## 4. Algorithm Selection

As the target variable is binary (True/False), a supervised classification algorithm is suitable. Various models were tested with different hyperparameters to identify the best one. Note: Logistic Regression is a classification algorithm, not regression in the traditional sense.

## 5. Evaluation

Models were evaluated using the following metrics:

• F1-score  
• ROC-AUC score  
• Confusion Matrix  
• Classification Report

## 6. Best Model

Logistic Regression with the following parameters:  
• penalty = l2  
• random\_state = 0  
• solver = newton-cg  
  
Performance:  
• F1-score ≈ 0.9916  
• ROC-AUC score = 1.0  
  
Conclusion: After comparing with other algorithms, Logistic Regression achieved the best performance. This model can be reliably used for predicting chronic kidney disease.

### Confusion Matrix

[[45, 0],  
 [ 1, 74]]

### Classification Report

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | Precision | Recall | F1-Score | Support |
| False | 0.98 | 1.00 | 0.99 | 45 |
| True | 1.00 | 0.99 | 0.99 | 75 |
| Accuracy |  |  | 0.99 | 126 |
| Macro Avg | 0.99 | 0.99 | 0.99 | 126 |
| Weighted Avg | 0.99 | 0.99 | 0.99 | 126 |