CHRONIC KIDNEY DISEASE PREDICTION

1.Problem statement:

To create best model which will predict the chronic kidney disease (CKD) based on the given dataset.

2.Dataset Information:

- Total Rows:399
- Total columns:25

3.Nominal data conversion:

Applied get_dummies () function to convert all the nominal data in the dataset to numbers.

4.Model comparison:

1.Logistic Regression:

Classification report:

	precision	recall	f1-score	support
False	0.98	1.00	0.99	45
True	1.00	0.99	0.99	75
accuracy			0.99	120
macro avg	0.99	0.99	0.99	120
weighted avg	0.99	0.99	0.99	120

Evaluation result:

Best parameters: {'penalty': 'l2', 'solver': 'lbfgs'}

Best F1_weighted score: 0.9916844

ROC_AUC value=1.0

2.SVC algorithm :

Classification report:

	precision	recall	f1-score	support
False	0.96	1.00	0.98	45
True	1.00	0.97	0.99	75
accuracy			0.98	120
macro avg	0.98	0.99	0.98	120
weighted avg	0.98	0.98	0.98	120

Evaluation result:

Best parameters: {'C': 10, 'gamma': 'auto', 'kernel': 'sigmoid'}

Best F1_weighted score: 0.983401

ROC_AUC value= 0.9997

3.***Decision Tree Classifier*:**

Classification report:

	precision	recall	f1-score	support
False	0.90	0.96	0.92	45
True	0.97	0.93	0.95	75
accuracy			0.94	120
macro avg	0.93	0.94	0.94	120
weighted avg	0.94	0.94	0.94	120

Evaluation metrics:

Best parameters: {'criterion': 'gini', 'max_features': 'sqrt', 'splitter': 'random'}

Best F1_weighted score: 0.94201228

ROC_AUC value= 0.944444

4.*RandomForestClassifier:*

Classification Report:

	precision	recall	f1-score	support
False	0.98	1.00	0.99	45
True	1.00	0.99	0.99	75
accuracy			0.99	120
macro avg	0.99	0.99	0.99	120
weighted avg	0.99	0.99	0.99	120

Evaluation metrics:

Best parameters: {'criterion': 'gini', 'max_features': 'log2', 'n_estimators': 50}

Best F1_weighted score: 0.99168449

ROC_AUC value= 0.999703

5.*KNN:*

Classification Report:

	precision	recall	f1-score	support
False	0.70	1.00	0.83	45
True	1.00	0.75	0.85	75
accuracy			0.84	120
macro avg	0.85	0.87	0.84	120
weighted avg	0.89	0.84	0.84	120

Evaluation metrics:

Best parameters: {'metric': 'manhattan', 'n_neighbors': 5, 'weights': 'uniform'}

Best F1_weighted score: 0.843984

ROC_AUC value= 0.953333

6.*NAIVE BAYES:*

GaussianNB:

Classification Report:

	precision	recall	f1-score	support
False	0.94	1.00	0.97	45
True	1.00	0.96	0.98	75
accuracy			0.97	120
macro avg	0.97	0.98	0.97	120
weighted avg	0.98	0.97	0.98	120

Evaluation metrics:

Best parameters: {'var_smoothing': np. float64(2.310129700083158e-09)}

Best F1_weighted score: 0.975148

ROC_AUC value= 1.0

MultinomialNB:

- Classification Report:

	precision	recall	f1-score	support
False	0.80	0.98	0.88	45
True	0.98	0.85	0.91	75
accuracy			0.90	120
macro avg	0.89	0.92	0.90	120
weighted avg	0.92	0.90	0.90	120

Evaluation metrics:

Best parameters: {'alpha': 0.001, 'fit_prior': True}

Best F1_weighted score: 0.901428

ROC_AUC value= 0.952592

BernoulliNB:

Classification Report:

	precision	recall	f1-score	support
False	0.94	1.00	0.97	45
True	1.00	0.96	0.98	75
accuracy			0.97	120
macro avg	0.97	0.98	0.97	120
weighted avg	0.98	0.97	0.98	120

[[45 0]

Evaluation metrics:

Best parameters: {'alpha': 0.001, 'fit_prior': True}

Best F1_weighted score: 0.9751481

ROC_AUC value= 0.996740

ComplementNB:

Classification Report:

	precision	recall	f1-score	support
False	0.79	0.98	0.87	45
True	0.98	0.84	0.91	75
accuracy			0.89	120
macro avg	0.89	0.91	0.89	120
weighted avg	0.91	0.89	0.89	120

Evaluation metrics:

Best parameters: {'alpha': 0.001, 'fit_prior': True}

Best F1_weighted score: 0.8932794

ROC_AUC value= 0.95259

Summary:

Based on the research the best model for predicting the chronic kidney disease is identified as Logistic regression with its ROC_AUC score as 1.0 and F1 weighted average score as 0.9916844.