ASSIGNMENT -CLASSIFICATION ALGORITHM

(Chronic Kidney Disease Prediction)

A) PROBLEM IDENTIFICATION:

1. DOMAIN SELECTION - MACHINE LEARNING
2. LEARNING SELECTION-SUPERVISED LEARNING
3. SUPERVISED LEARNING -CLASSIFICATION Q

B)BASIC INFORMAION:

Data set has 399 Rows and 25 Columns.

Total Counts:

1 249

0 150

C)PRE PROCESSING METHOD:

Data set has 14 ordinal data columns.so we converted ordinal data to numerical data by using one hot encoding.

D) BEST MODEL:

BERNOULLI -NAVIE BAYES

- ACCURACY :0.99
- F1 MACRO :0.99
- ROC_AUC_SCORE:1.0
- BEST CONFUSION MATRIX RESULT

```
confusion matrix:
 [[51  0]
 [ 1 81]]
```

E) MODELS RESULT:

- Logistic Regression:

Classification Report:

	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

```
Confusion Matrix:
 [[51  0]
 [ 1 81]]
```

- The f1 macro value for best parameter:{'penalty': 'l2', 'solver ': 'lbfgs'} = 0.9924946382275899
- roc_auc_score: 1.0

❖ Decision tree:

Classification Report:

	precision	recall	f1-score	support
0	0.93	1.00	0.96	51
1	1.00	0.95	0.97	82
accuracy			0.97	133
macro avg	0.96	0.98	0.97	133
weighted avg	0.97	0.97	0.97	133

```
confusion matrix:
 [[51  0]
 [ 4 78]]
```

- The f1 macro value for best parameter {'criterion': 'gini', 'maxfeatures': 'log2', 'splitter': 'random'} = 0.9701163285572423
- Roc_auc_score : 0.97

❖ RandomForest:

Classification Report:

	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

```
confusion matrix:
 [[50  1]
 [ 1 81]]
```

- The f1 macro value for best parameter {'criterion': 'entropy', 'max_features': 'sqrt', 'n_estimators': 100}: 0.9849624060150376
- roc_auc_score: 0.99

❖ KNN :

Classification Report:

	precision	recall	f1-score	support
0	0.89	1.00	0.94	51
1	1.00	0.93	0.96	82
accuracy			0.95	133
macro avg	0.95	0.96	0.95	133
weighted avg	0.96	0.95	0.96	133

```
confusion matrix:
 [[51  0]
 [ 6 76]]
```

- The f1 macro value for best parameter {'algorithm': 'auto', 'leaf_size': 30, 'metric': 'minkowski', 'n_neighbors': 5, 'p': 2, 'weights': 'distance'}: 0.955283779067923

- roc_auc_score: 1.0

❖ SVM:

Classification Report:

	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

```
Confusion Matrix:
 [[51  0]
 [ 2 80]]
```

- The f1 macro value for best parameter:{'C': 10, 'kernel': 'sigmoid'} 0.9850141736106648

❖ roc_auc_score: 1.0

❖ BERNOULLINB:

Classification Report:

	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

```
confusion matrix:
 [[51  0]
 [ 1 81]]
```

- The f1 macro value for best parameter BernoulliNB(): 0.9924946382275899
- roc_auc_score: 1.0

❖ GAUSSIANNB:
Classification Report:

	precision	recall	f1-score	support
0	0.94	1.00	0.97	51
1	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

```
confusion matrix:
 [[51  0]
 [ 3 79]]
```

- The f1 macro value for best parameter GaussianNB(): 0.9775556904684072
- roc_auc_score: 1.0