Chronic Kidney Disease Prediction Model

1. Problem Statement :

   Client has provided a dataset with multiple patient's medical history data with various parameters.The requirement is to predict whether the patient is susceptible for kidney disease or not.

2. Dataset :

   Columns : Parameters like Age ,BP ,RBC ., etc (25 Columns)

   Rows : Number of Patients - 399

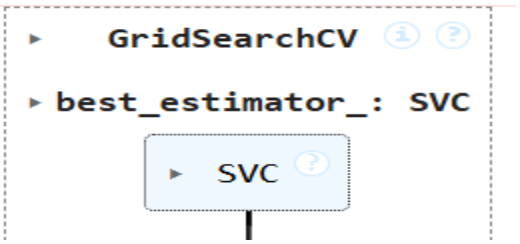
3. Data Pre-Processing Method :

   Columns with Categorical (Qualitative) is converted to Nominal data & Standard Scaler method.

4. Model results :

LogisticRegression

	precision	recall	f1-score	support
0	0.92	0.94	0.93	51
1	0.96	0.95	0.96	82
accuracy			0.95	133
macro avg	0.94	0.95	0.94	133
weighted avg	0.95	0.95	0.95	133



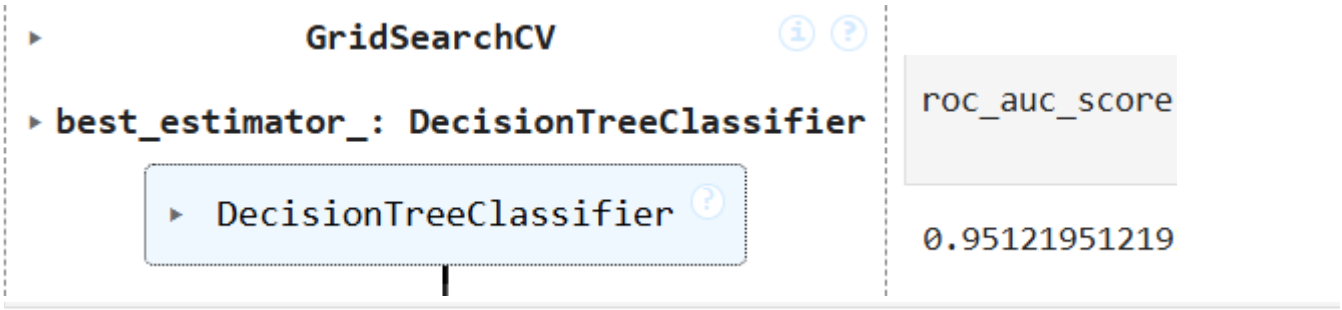

```
roc_auc_score(y_tes

0.9985652797704447
```

The clf_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

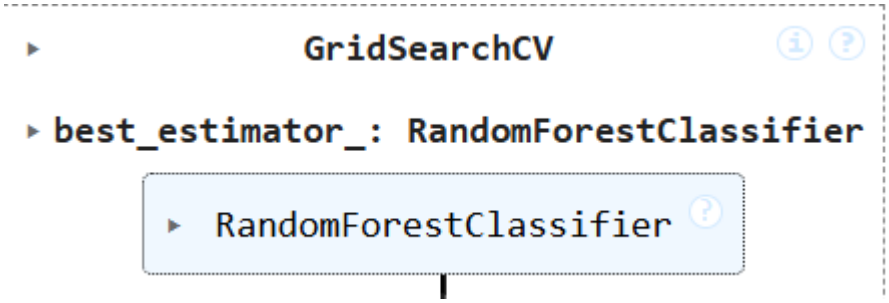
The f1_macro value of the best parameter{'C': 10.0, 'gamma': 'auto', 'kernel': 'sigmoid'} : 0.9701163285572423



The f1_macro value for best parameter {'criterion': 'entropy', 'max_features': 'sqrt', 'splitter': 'random'}: 0.940494593126172

The report:

	precision	recall	f1-score	support
0	0.86	1.00	0.93	51
1	1.00	0.90	0.95	82
accuracy			0.94	133
macro avg	0.93	0.95	0.94	133
weighted avg	0.95	0.94	0.94	133



The f1_macro value for best parameter {'criterion': 'gini', 'max_features': 'sqrt', 'n_estimators': 100}: 0.9849624060150376

The 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

Final result :

Random Forest Classification model seems to be the best prediction model as it has the highest f1 score & Area under the curve (roc) values.