**Chronic Kidney Disease Prediction**

**Aim:** To find the best model using machine learning algorithms.

***1.Identify your problem statement***

Domain Selection: Machine Learning

Learning Selection: Supervised Learning

Output-Numerical value: Classification

***2.Dataset Information***

Number of inputs: 'age', 'bp', 'al', 'su', 'bgr', 'bu', 'sc', 'sod', 'pot', 'hrmo', 'pcv', 'wc', 'rc', 'sg\_b', 'sg\_c', 'sg\_d', 'sg\_e', 'rbc\_normal', 'pc\_normal', 'pcc\_present', 'ba\_present', 'htn\_yes', 'dm\_yes', 'cad\_yes', 'appet\_yes', 'pe\_yes', 'ane\_yes'

Output: classification\_yes

Value counts: classification\_yes

1🡪🡪249

0🡪🡪150

Total number of rows: 399

Total number of columns: 27

***3. Pre-processing method***

Few inputs were provided under categorical values as nominal data which has been converted to numerical values using ‘one-hot-encoding’.

Standardised the values in order to create a best model.

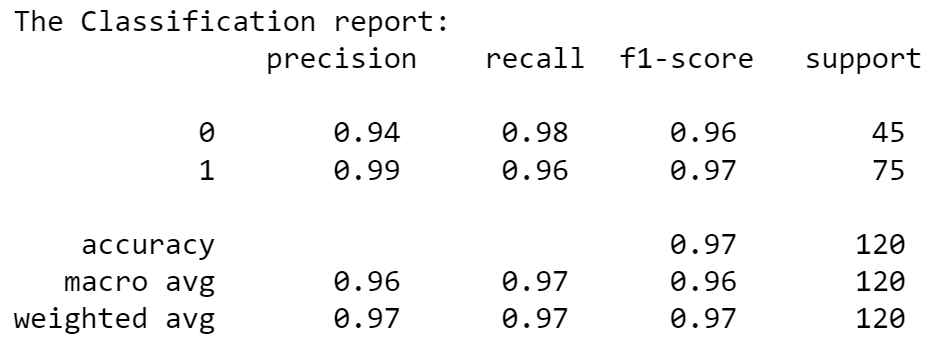
***4. Develop a good model with evaluation metrics***

Performed the below mentioned machine learning algorithms to find the best model using Grid Search Cross validation.

Decision Tree

roc\_auc\_score: 0.9688888888888889

The f1\_macro value for best parameter {'criterion': 'log\_loss', 'max\_features': 'log2', 'splitter': 'random'}: 0.9668037602820211



Random Forest

roc\_auc\_score: 0.9997037037037038

The f1\_macro value for best parameter {'criterion': 'gini'}: 0.9833333333333335

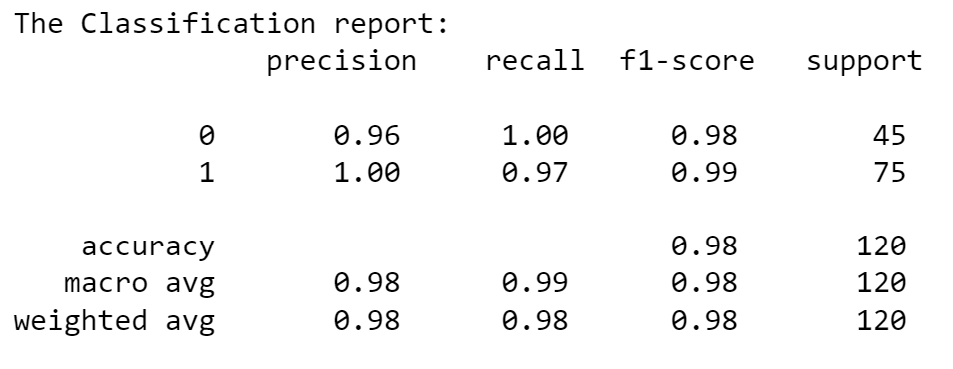
A number of numbers on a white background

Description automatically generated

Support Vector Machine

roc\_auc\_score: 0.9997037037037036

The f1\_macro value for best parameter {'C': 10, 'gamma': 'scale', 'kernel': 'sigmoid'}: 0.9834018801410106



Logistic Regressor

roc\_auc\_score: 1.0

The f1\_macro value for best parameter {'penalty': 'l2', 'solver': 'lbfgs'}: 0.9916844900066377

A screenshot of a graph

Description automatically generated

KNN

roc\_auc\_score: 1.0

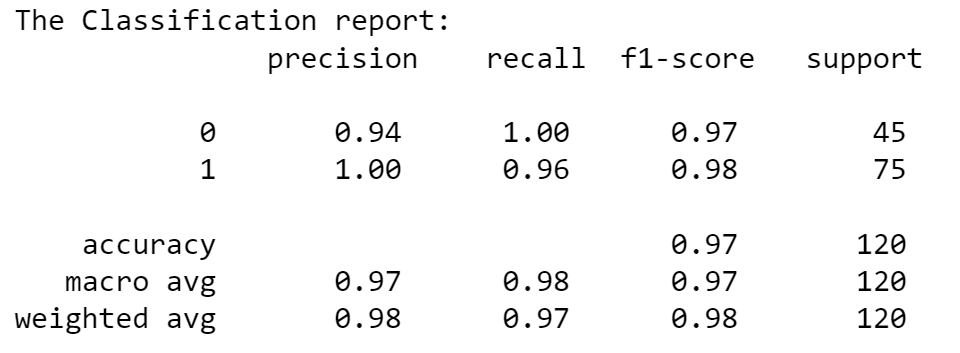
The f1\_macro value for best parameter {'algorithm': 'auto', 'weights': 'distance'}: 0.9505208333333334

A number of numbers on a white background

Description automatically generated

Naïve Bayes

The f1\_macro value: 0.9751481237656352



Result: It is concluded that the LOGISTIC REGRESSOR with penalty=l2, solver=lbfgs provides the best model.