

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

The Classification report:

	precision	recall	f1-score	support
0	0.94	0.98	0.96	45
1	0.99	0.96	0.97	75
accuracy			0.97	120
macro avg	0.96	0.97	0.96	120
weighted avg	0.97	0.97	0.97	120

RANDOM FOREST

ROC_AUC_SCORE: 0.9997037037037038

The f1_macro value for best parameter {'criterion': 'gini'}: 0.9833333333333335

The Classification report:

	precision	recall	f1-score	support
0	0.98	0.98	0.98	45
1	0.99	0.99	0.99	75
accuracy			0.98	120
macro avg	0.98	0.98	0.98	120
weighted avg	0.98	0.98	0.98	120

SUPPORT VECTOR MACHINE

ROC_AUC_SCORE: 0.9997037037037036

The f1_macro value for best parameter {'C': 10, 'gamma': 'scale', 'kernel': 'sigmoid'}:
0.9834018801410106

The Classification report:

	precision	recall	f1-score	support
0	0.96	1.00	0.98	45
1	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

LOGISTIC REGRESSOR

ROC_AUC_SCORE: 1.0

The f1_macro value for best parameter {'penalty': 'l2', 'solver': 'lbfgs'}: 0.9916844900066377

The Classification report:

	precision	recall	f1-score	support
0	0.98	1.00	0.99	45
1	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

KNN

ROC_AUC_SCORE: 1.0

The f1_macro value for best parameter {'algorithm': 'auto', 'weights': 'distance'}: 0.9505208
333333334

The Classification report:

	precision	recall	f1-score	support
0	0.88	1.00	0.94	45
1	1.00	0.92	0.96	75
accuracy			0.95	120
macro avg	0.94	0.96	0.95	120
weighted avg	0.96	0.95	0.95	120

NAÏVE BAYES

The f1_macro value: 0.9751481237656352

The Classification report:

	precision	recall	f1-score	support
0	0.94	1.00	0.97	45
1	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

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