

MACHINE LEARNING CLASSIFICATION

ASSIGNMENT STATEMENT

1. PROBLEM IDENTIFICATION:

- First, we collect the data from the client, then from the collected data, we have to predict if the patient is affected by Chronic Kidney Disease (CKD) as yes or no.
- Machine learning → Supervised learning → Classification

2. Info about the dataset:

- 399 rows × 25 columns

3. Pre-processing method:

- pc_normal(normal & abnormal), pcc_present(present & not present), ba_present(present & not present), htn_yes(yes & no), dm_yes(yes & no), cad_yes(yes & no), appet_yes(yes & no), pe_yes(yes & no), ane_yes(yes & no), classification_yes(yes & no).

4. Model :

- Random Forest

5. Research values:

- SVM

The 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

- RF

The 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

- Naïve bayes

The 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

- Logistic regression

The 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

- KNN

	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

- DT

The confusion Matrix:

	precision	recall	f1-score	support
0	0.94	0.98	0.96	51
1	0.99	0.96	0.98	82
accuracy			0.97	133
macro avg	0.97	0.97	0.97	133
weighted avg	0.97	0.97	0.97	133

6. FINAL MODEL NAME:

- Random Forest – This is a good model that's confusion_matrix(accuracy) value is near by 1 with comparing other models.