

## PROBLEM STATEMENT:

A requirement from the Hospital, Management asked us to create a predictive model which will predict the Chronic Kidney Disease (CKD) based on the several parameters. The Client has provided the dataset of the same.

Prediction: whether the patient has kidney disease or not?

- Domain Selection: Machine Learning
- Learning Selection: Supervised Learning (Requirement is clear)
- Classification(label is in categorical values)

Information about the given dataset:

Total no.of.rows = 399

Total no.of.columns = 28

- I have done preprocessing method in the columns “sg, rbc, pc, pcc, ba ,htn ,dm, cad, appet, pe, ane” (converting nominal data to numerical data using (one hot encoding function )

## 1.SVM algorithm:

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

Roc\_auc\_score:

0.9997037037037036

## 2.Decision tree algorithm:

Classification report :

	precision	recall	f1-score	support
0	0.89	0.93	0.91	45
1	0.96	0.93	0.95	75
accuracy			0.93	120
macro avg	0.93	0.93	0.93	120
weighted avg	0.93	0.93	0.93	120

---

Roc\_auc\_score:

0.9333333333333333

### 3.Random Forest Algorithm:

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

Roc\_auc\_score:

0.9994074074074075

### 4.Logistic regression:

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

Roc\_auc\_score:

1.0

### 5.K nearest neighbors algorithm:

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

Roc\_auc\_score:

0.9995555555555555

Naïve bayes algorithm:

6.complement NB:

Classification report:

---

	precision	recall	f1-score	support
0	0.79	0.98	0.87	45
1	0.98	0.84	0.91	75
accuracy			0.89	120
macro avg	0.89	0.91	0.89	120
weighted avg	0.91	0.89	0.89	120

Roc\_auc\_score:

0.9525925925925927

7.gaussian NB:

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

Roc\_auc\_score:

1.0

8.bernoulli NB:

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

Roc\_auc\_score:

1.0

9.multinomial NB:

Classification report:

	precision	recall	f1-score	support
0	0.80	0.98	0.88	45
1	0.98	0.85	0.91	75
accuracy			0.90	120
macro avg	0.89	0.92	0.90	120
weighted avg	0.92	0.90	0.90	120

Roc\_auc\_score:

0.9525925925925927

10.categorical NB:

Classification report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	150
1	1.00	1.00	1.00	249
accuracy			1.00	399
macro avg	1.00	1.00	1.00	399
weighted avg	1.00	1.00	1.00	399

Roc\_auc\_score:

1.0

- Using the above 10 algorithms, we got the best model in **logistic regression** algorithm (**accuracy=0.99,roc\_auc value=1.0**).so we can proceed to the next phase ((i.e)deployment phase) .
- The best parameters are: {'penalty': 'l2', 'solver': 'lbfgs'}