

Problem Statement or Requirement: 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.

1.) Identify your problem statement :

Based on the given dataset we need to predict the Chronic Kidney disease based on the given parameters in the dataset

2.) Tell basic info about the dataset (Total number of rows, columns)

Rows:399

Columns:25

3.) Mention the pre-processing method if you're doing any (like converting string to number – nominal data)

The preprocessing is done for the below columns:

rbc,pc,pcc,ba,htn,dm,cad,appet,pe,ane,classification

4.) Develop a good model with good evaluation metric. You can use any machine learning algorithm; you can create many models. Finally, you have to come up with final model.

The final model or Best model is **Random forest** with F1 score of **0.9916844900066377**

Best Parameter:{'criterion': 'gini', 'n_estimators': 100}

5.) All the research values of each algorithm should be documented. (You can make tabulation or screenshot of the results.)

Random Forest:

The classification report is

	precision	recall	f1-score	support
False	0.98	1.00	0.99	45
True	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:0.9997037037037038

Logistic Regression:

The report is:

	precision	recall	f1-score	support
False	0.96	0.96	0.96	45
True	0.97	0.97	0.97	75
accuracy			0.97	120
macro avg	0.96	0.96	0.96	120
weighted avg	0.97	0.97	0.97	120

ROC_AUC_SCORE:0.997925925925926

SVC:

The report is:

	precision	recall	f1-score	support
False	0.78	0.93	0.85	45
True	0.95	0.84	0.89	75
accuracy			0.88	120
macro avg	0.87	0.89	0.87	120
weighted avg	0.89	0.88	0.88	120

ROC_AUC_Score:0.9265185185185185

Naive Bayes:

1)GaussianNB

The report is:

	precision	recall	f1-score	support
False	0.96	1.00	0.98	45
True	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:1.0

2)BernoulliNB

The confusion matrix is:

```
[[43  2]
 [20 55]]
```

The report is:

	precision	recall	f1-score	support
False	0.68	0.96	0.80	45
True	0.96	0.73	0.83	75
accuracy			0.82	120
macro avg	0.82	0.84	0.81	120
weighted avg	0.86	0.82	0.82	120

ROC_AUC_Score:0.845037037037037

3) MultinomialNB

Fitting 5 folds for each of 3 candidates, totalling 15 fits

	precision	recall	f1-score	support
False	0.67	0.98	0.79	45
True	0.98	0.71	0.82	75
accuracy			0.81	120
macro avg	0.82	0.84	0.81	120
weighted avg	0.86	0.81	0.81	120

ROC_AUC_Score:0.9099259259259259

4) CategoricalNB

The confusion matrix is:

```
[[44  1]
 [22 53]]
```

The report is:

	precision	recall	f1-score	support
False	0.67	0.98	0.79	45
True	0.98	0.71	0.82	75
accuracy			0.81	120
macro avg	0.82	0.84	0.81	120
weighted avg	0.86	0.81	0.81	120

ROC_AUC_Score:

5) ComplementNB

The confusion matrix is:

```
[[44  1]
 [22 53]]
```

The report is:

	precision	recall	f1-score	support
False	0.67	0.98	0.79	45
True	0.98	0.71	0.82	75
accuracy			0.81	120
macro avg	0.82	0.84	0.81	120
weighted avg	0.86	0.81	0.81	120

ROC_AUC_Score:0.9188148148148149

Decision Tree:

The classification report is

	precision	recall	f1-score	support
False	0.96	1.00	0.98	45
True	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:0.9866666666666667

KNN

The confusion matrix is

```
[[41 10]
 [22 60]]
```

The classification report is

	precision	recall	f1-score	support
False	0.65	0.80	0.72	51
True	0.86	0.73	0.79	82
accuracy			0.76	133
macro avg	0.75	0.77	0.75	133
weighted avg	0.78	0.76	0.76	133

roc_auc_score:0.7678144428503109