

Classification Assignment

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

Machine Learning → Supervised Learning → Regression

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

Input -age,

bp,sg,al,su,rbc,pc,pcc,ba,bgr,bu,sc,sod,pot,hrmo,pcv,wc,rc,htn,dm,cad,appet,pe,ane

Output - classification

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

We need to convert character to numbers. We can achieve that by using Pandas library. We have some of the columns with nominal data. Convert character to numbers by using `get_dummies`

sg, 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.

Decision Tree – 1.0
Random forest -0.94
SVM – 0.9924
Logistic – 0.9924
KNN- 0.9440
Naves Bayes -0.9800
AdaBoost – 0.9775
XgBoost - 0.9774
LGBost - 1.0
Extra Classification - 0.9924

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

SVM :

	precision	recall	f1-score	support
0	0.94	1.00	0.97	51
1	1.00	0.96	0.98	82
accuracy			0.98	133
macro avg	0.97	0.98	0.98	133
weighted avg	0.98	0.98	0.98	133

Decision Tree:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	51
1	1.00	1.00	1.00	82
accuracy			1.00	133
macro avg	1.00	1.00	1.00	133
weighted avg	1.00	1.00	1.00	133

Random Forest:

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

Logistic Classification:

	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

Extra tree Classification:

	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

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

Naïve Bayes:

	precision	recall	f1-score	support
0	0.94	1.00	0.97	51
1	1.00	0.96	0.98	82
accuracy			0.98	133
macro avg	0.97	0.98	0.98	133
weighted avg	0.98	0.98	0.98	133

```
[[51  0]
 [ 3 79]]
```

Adaboost Algorithm:

	precision	recall	f1-score	support
0	0.94	1.00	0.97	51
1	1.00	0.96	0.98	82
accuracy			0.98	133
macro avg	0.97	0.98	0.98	133
weighted avg	0.98	0.98	0.98	133

XGBoost:

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

LightGBM:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	51
1	1.00	1.00	1.00	82
accuracy			1.00	133
macro avg	1.00	1.00	1.00	133
weighted avg	1.00	1.00	1.00	133

6. Mention your final model, justify why u have chosen the same.

My Justify model for this dataset is SVM and Decision Tree.

Even though Decision tree is showing 100% accuracy, We are choosing SVM over Decision Tree. Since it is Unstable.

Base on the client requirement, We need to take action on other models.