

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

classification

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

25 columns and 400 rows

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

Below are the pre-processing column to convert string to nominal data

rbc_normal', 'pc_normal', 'pcc_present', 'ba_present', 'htn_yes', 'dm_yes', 'cad_yes',
'appet_yes', 'pe_yes', 'ane_yes', 'classification_yes'

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.

SVM grid is the final model.

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

File Edit View Insert Cell Kernel Widgets Help

Save + Undo Copy Paste Up Down Run Stop Restart Code

```
In [15]: print("The report:\n",clf_report)
```

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

```
In [16]: from sklearn.metrics import roc_auc_score
roc_auc_score(y_test,grid.predict_proba(X_test)[:,:1])
```

Out[16]: 1.0

File Edit View Insert Cell Kernel Widgets Help

Save + Undo Copy Paste Up Down Run Stop Restart Code

[0 82]

```
In [14]: print("The report:\n",clf_report)
```

The report:

	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

```
In [15]: from sklearn.metrics import roc_auc_score
roc_auc_score(y_test,grid.predict_proba(X_test)[:,:1])
```

Out[15]: 1.0

[8 / 4]

In [14]: `print("The report:\n",clf_report)`

The report:

	precision	recall	f1-score	support
0	0.86	0.96	0.91	51
1	0.97	0.90	0.94	82
accuracy			0.92	133
macro avg	0.92	0.93	0.92	133
weighted avg	0.93	0.92	0.93	133

In [15]: `from sklearn.metrics import roc_auc_score`

`roc_auc_score(y_test,grid.predict_proba(X_test)[:,:1])`

Out[15]: 0.9316116690578671

```
In [14]: print("The report:\n",clf_report)
```

The report:

	precision	recall	f1-score	support
0	1.00	0.98	0.99	51
1	0.99	1.00	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

```
In [15]: from sklearn.metrics import roc_auc_score
roc_auc_score(y_test,grid.predict_proba(X_test)[:,:1])
```

Out[15]: 1.0

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

SVM grid is the final model. Because most of the parameters are 1.0.