

Classification Assignment Problem

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

To predict the Chronic Kidney disease based on the inputs provided by the Hospital Management

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

399 rows × 25 columns

`classification`

`yes 249`

`no 150`

`Name: count, dtype: int64`

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

The One hot Encoding method is used to convert categorical columns into ordinal data

After 399 rows * 40 columns

Standardization to improve the model as there are many rows

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.

Random Forest makes best model with accuracy of 0.99

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

In [11]: Algorithms.SVM()

```
[[[48  3]
   [ 2 80]]

  [[80  2]
   [ 3 48]]]
      precision    recall  f1-score   support

      0       0.96      0.98      0.97        82
      1       0.96      0.94      0.95        51

   accuracy          0.96          133
  macro avg          0.96      0.96      0.96          133
 weighted avg          0.96      0.96      0.96          133
```

In [12]: Algorithms.DT()

```
[[[48  3]
   [ 2 80]]

  [[80  2]
   [ 3 48]]]
      precision    recall  f1-score   support

      0       0.96      0.98      0.97        82
      1       0.96      0.94      0.95        51

   accuracy          0.96          133
  macro avg          0.96      0.96      0.96          133
 weighted avg          0.96      0.96      0.96          133
```

In [13]: Algorithms.RF()

```
[[[51  0]
   [ 1 81]]

  [[81  1]
   [ 0 51]]]
      precision    recall  f1-score   support

      0       1.00      0.99      0.99        82
      1       0.98      1.00      0.99        51

   accuracy          0.99          133
  macro avg          0.99      0.99      0.99          133
 weighted avg          0.99      0.99      0.99          133
```

C:\Users\Dharani Vinoth\anaconda3\Lib\site-packages\sklearn\base.py:1151: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
return fit_method(estimator, *args, **kwargs)

In [14]: Algorithms.KNN()

```
[[[51  0]
   [ 7 75]]

  [[75  7]
   [ 0 51]]]
      precision    recall  f1-score   support

      0       1.00      0.91      0.96        82
      1       0.88      1.00      0.94        51

   accuracy          0.95          133
  macro avg          0.94      0.96      0.95          133
 weighted avg          0.95      0.95      0.95          133
```

```
In [15]: Algorithms.LGR()
```

```
[[[51  0]
   [ 2 80]]
```

```
[[[80  2]
   [ 0 51]]]
```

	precision	recall	f1-score	support
0	1.00	0.98	0.99	82
1	0.96	1.00	0.98	51
accuracy			0.98	133
macro avg	0.98	0.99	0.98	133
weighted avg	0.99	0.98	0.99	133

```
C:\Users\Dharani Vinoth\anaconda3\Lib\site-packages\sklearn\utils\validation
was passed when a 1d array was expected. Please change the shape of y to (n
y = column_or_1d(y, warn=True)
```