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 :

a)Supervised Learning:

Input and Output should be present and Client’s requirement should be very clear.

b) Machine Learning:

Client has provided the dataset in names and numbers

c)Classification:

Output is a categorical data so we should say Classification.

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

Total number of rows is 399

Total number of columns is 25

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

Converting Categorical Data to Numerical.

Nominal data used Algorithm is One Hot Encoding and Column can be Expanded

1. 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.

1.SVM-Classification:

'C':10,'gamma':'auto','kernel':'poly'}: 1.0

2.Decision Tree-Classification:

The f1\_macro value for best parameter

'criterion': 'entropy',

'max \_ features': 'sqrt'

‘splitter': 'random' : 0.9550425208319945

3.Random Forest-Classification:

The f1\_macro value for best parameter

'criterion': 'entropy',

'max \_ features': 'log2',

'n \_ estimators': 100 : 0.9850064683509052

4. Logistic Regression-Classification:

The f1\_macro value for best parameter

'penalty': 'l2',

'solver': 'newton-cg': 0.9924946382275899

5. KKN-Classification:

The f1\_macro value for best parameter

'algorithm': 'auto',

'metric': 'minkowski',

'n\_neighbors': 5,

'p': 2,

'weights': 'distance' : 0.7158175158175157

6. Naive Bayes:

The f1\_macro value for best parameter : 0.9700129367018107

5.)  All the research values of each algorithm should

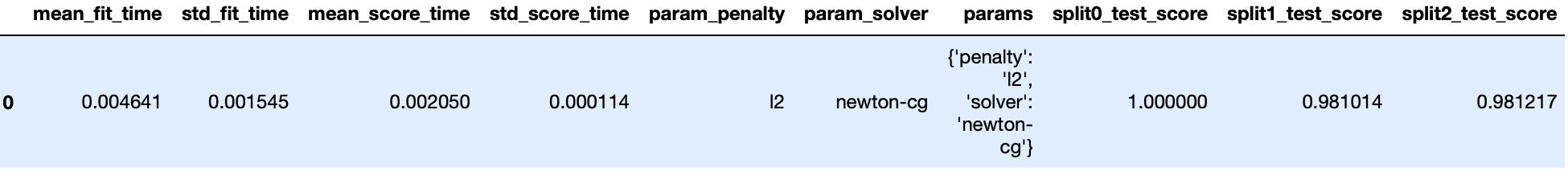
be documented

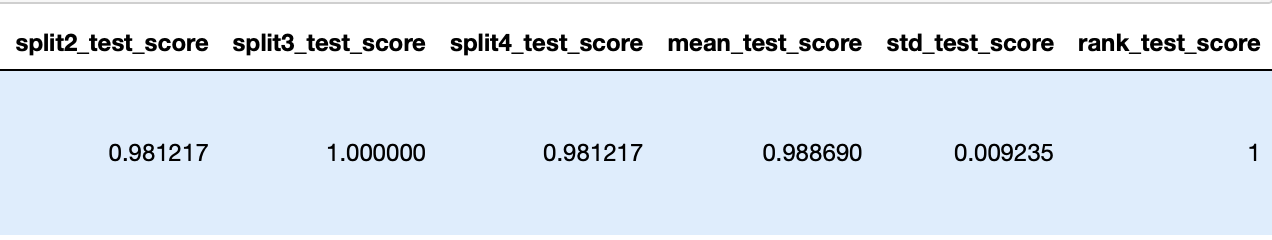
(You can make tabulation or screenshot of the results.)

|  |  |
| --- | --- |
| Algorithm | Average accuracy |
| SVM-Classification | 1.0 |
| Decision Tree-Classification | 0.95 |
| Random Forest-Classification | 0.98 |
| Logistic Regression-Classification | 0.99 |
| KKN-Classification | 0.71 |
| Naive Bayes | 0.97 |

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

The final machine learning best method of classification: Logistic Regression-Classification confusion matrix value is 0.99





Classification Assignment:

Git hub link in SVM-Classification:

<https://github.com/Vidhyavino/Machine-Learning-Classification-Assignment/blob/f2ef8a8b1e500f5f11ba6f2f642f1496473e0d57/%20SVC-Classification.ipynb>

Git hub link in Decision Tree-Classification:

<https://github.com/Vidhyavino/Machine-Learning-Classification-Assignment/blob/f2ef8a8b1e500f5f11ba6f2f642f1496473e0d57/%20Decision%20Tree-Classification.ipynb>

Git hub link in Random Forest-Classification:

<https://github.com/Vidhyavino/Machine-Learning-Classification-Assignment/blob/4674d8533d7c1afebf4a8eb78b19ff0a43f65833/%20Random%20forest-classification.ipynb>

Git hub link in Logistic Regression-Classification:

<https://github.com/Vidhyavino/Machine-Learning-Classification-Assignment/blob/78553ffa68a7c898faaea434220ed68e3e1b6de5/%20Logistic%20Regression%20-%20classification.ipynb>

Git hub link in KKN-Classification:

<https://github.com/Vidhyavino/Machine-Learning-Classification-Assignment/blob/217adc574714c7e5dc9f4a1f97aeaf9768425662/%20KNN-Classification.ipynb>

Git hub link in Naïve Bayes: <https://github.com/Vidhyavino/Machine-Learning-Classification-Assignment/blob/8aa76b7d84afd58fac5cd900d7ae2467a64af578/%20Naive%20bayes-%20classification.ipynb>