PROJECT : PREDICTING CHRONIC KIDNEY DISEASE

3 stages of selection

1) Stage1 - Machine Learning Domain
2) Stage2 - Supervised Learning
3) Stage3 - Classification

Project details

1) Problem statement : Predicting Chronic Kidney Disease

2) Basic information about dataset

- No of rows in datasat - 399
- No of columns in dataset - 28
- Input / Independent data - (27No's) 'age', 'bp', 'al', 'su', 'bgr', 'bu', 'sc', 'sod', 'pot', 'hrmo', 'pcv', 'wc', 'rc', 'sg_b', 'sg_c', 'sg_d', 'sg_e', 'rbc_normal', 'pc_normal', 'pcc_present', 'ba_present', 'htn_yes', 'dm_yes', 'cad_yes', 'appet_yes', 'pe_yes', 'ane_yes'
- Output / Dependent data - classification_yes

3) Preprocessing method - Label encoder to convert ordinal categorical data into numerical data

4) Machine learning algorithms used

A. Support Vector Machine
B. Decision Tree
C. Random Forest

A) Algorithm : Support Vector Machine

```
Confusion Matrix:
 [[51  0]
 [ 1 81]]
```

```
classification_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
```

```
The f1_macro value for best parameter{'C': 10, 'gamma': 'auto', 'kernel': 'sigmoid'}:
 0.9924946382275899
```

```
roc_auc_score= 1.0
```

B)Algorithm : Decision Tree

```
Confusion Matrix:
 [[51  0]
 [ 4 78]]
```

```
classification_report:
              precision    recall  f1-score   support

           0       0.93      1.00      0.96        51
           1       1.00      0.95      0.97        82

    accuracy                           0.97       133
   macro avg       0.96      0.98      0.97       133
weighted avg       0.97      0.97      0.97       133
```

```
The f1_macro value for best parameter{'criterion': 'entropy', 'max_features': 'auto', 'splitter': 'random'}:
 0.9701163285572423
```

```
roc_auc_score= 0.975609756097561
```

C) Algorithm : Random forest

```
Confusion Matrix:
 [[50  1]
 [ 1 81]]
```

```
classification_report:
              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
```

```
The f1_macro value for best parameter{'criterion': 'entropy', 'max_features': 'log2', 'n
 0.9849624060150376
```

```
roc_auc_score= 0.9997608799617408
```

5) Conclusion

a)Best Algorithm for given problem statement: Support Vector Machine

b)Type1 error < Type2 error (0<1)

c)f1 macro value : 0.99

d)Roc-auc_score : 1

e)Accuracy : 0.99

Finalized Result

Support Vector Machine Algorithm can be used with following hyper tuning parameters

Parameters : C=10,
gamma: auto,
kernel: sigmoid