

Classification assignment

31/08/2023

1.PROBLEM STATEMENT

The given ckd dataset is machine learning classification algorithm

2.given dataset is 399 rows × 25 columns

Develop a various model

1.Logistic regression classification (grid search):

Classification report of logistic regression

```
from sklearn.metrics import classification_report
clf_report = classification_report(y_test, grid_predictions)
print(clf_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

F1 macro value:

The f1_macro value for best parameter {'penalty': 'l2', 'solver': 'newton-cg'}: 0.9924946382275899

F1 macro value of logistic regression=0.9924

Roc_auc score :

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

: 1.0

Roc auc score of logistic regression=1

logistic regression is best model of CKD.CSV dataset because of roc auc score is 1 compared to other classification algorithm of machine learning

DECISION TREE CLASSIFICATION GRID

CLASISIFICATION REPORT:

```
print(clf_report)
```

	precision	recall	f1-score	support
0	0.93	0.98	0.95	51
1	0.99	0.95	0.97	82
accuracy			0.96	133
macro avg	0.96	0.97	0.96	133
weighted avg	0.96	0.96	0.96	133

F1 Macro value for best parameter:

```
The f1_macro value for best parameter {'criterion': 'entropy', 'max_features': 'sqrt', 'splitter': 'random'}: 0.9625928174473452
```

F1 score of decision tree=0.9625

Roc-auc score:

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

0.9658058345289334

Roc auc score of decision tree=0.9658

3.Random forest classification grid:

Classification report:

```
print(clf_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

F1 macro score

```
from sklearn.metrics import f1_score
f1_macro=f1_score(y_test,grid_predictions,average='weighted')
print("The f1_macro value for best parameter {}".format(grid.best_params_),f1_macro)
```

The f1_macro value for best parameter {'criterion': 'gini', 'max_features': 'log2', 'n_estimators': 10}: 0.9849624060150376

F1 macro score=0.984

Roc auc:

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

0.8478000956480153

Roc auc score of random forest=0.847(it poor model compared to all other classification algorithm of this dataset

4. svm of classification grid

Classification report

```
print(clf_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

F1 macro score

```
from sklearn.metrics import f1_score
f1_macro=f1_score(y_test,grid_predictions,average='weighted')
print("The f1_macro value for best parameter {}".format(grid.best_params_),f1_macro)
```

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

F1 macro score=0.992 and

best parameter:c=10,gamma=auto,kernel=sigmoid

CONCLUSION:

BEST MODEL=LOGISTIC REGRESSION

Roc auc score of logistic regression=1

it is a best model of CKD.CSV dataset because of roc auc score is 1 compared to other classification algorithm of machine learning