PROJECT DEVELOPMENT PHASE

MODEL PERFORMANCE TEST

Date	17 Nov 2022
Team ID	PNT2022TMID37140
Project Name	Early Detection of Chronic Kidney Disease Using Machine Learning
Maximum Marks	10 Marks

Model Performace Testing:

Project team shall fill the following information in model performance testing template.

S.no	parameters	Values	screenshot
1	Metrics	**Regression Model:** MAE - , MSE - , RMSE - , R2 score - **Classification Model:** Confusion Matrix - , Accuray Score- & Classification Report -	See below
2	Tune the model	Hyperparameter Tuning - Validation Method -	See below

1. Metrics

Model: Decision Tree Classification

```python
# accuracy score, confusion matrix and classification report of decision tree

dtc_acc = accuracy_score(y_test, dtc.predict(X_test))

print(f"Training Accuracy of Decision Tree Classifier is {accuracy_score(y_train, dtc.predict(X_train))}")
print(f"Test Accuracy of Decision Tree Classifier is {dtc_acc} \n")

print(f"Confusion Matrix :- \n{confusion_matrix(y_test, dtc.predict(X_test))}\n")
print(f"Classification Report :- \n {classification_report(y_test, dtc.predict(X_test))}")
```

```
Training Accuracy of Decision Tree Classifier is 1.0
Test Accuracy of Decision Tree Classifier is 0.9333333333333333

Confusion Matrix :-
[[70  2]
 [ 6 42]]

Classification Report :-
              precision    recall  f1-score   support

           0       0.92      0.97      0.95        72
           1       0.95      0.88      0.91        48

    accuracy                           0.93       120
   macro avg       0.94      0.92      0.93       120
weighted avg       0.93      0.93      0.93       120
```

2. Tune the Model

Hyperparameter Tuning:

- The number of features is important and should be tuned in decision tree classification.
- Initially all parameters in the dataset are taken as independent values to arrive at the dependent decision of Chronic Kidney Disease or No Chronic Kidney Disease.
- But the result was not accurate so used only 8 more correlated values as independent values to arrive at the dependent decision of Chronic Kidney Disease or not.

Validation Method:

It involves **partitioning the training data set into subsets, where one subset is held out to test the performance of the model**. This data set is called the validation data set.

Cross validation is to use different models and identify the best:

Logistic Regression Model performance values:

```
accuracy_score(y_test,y_pred)
```

```
0.925
```

```
conf_mat=confusion_matrix(y_test,y_pred)
conf_mat
```

```
array([[48,  6],
       [ 0, 26]], dtype=int64)
```

```
print(classification_report(y_test,y_pred))
```

```
              precision    recall  f1-score   support

           0       1.00      0.89      0.94        54
           1       0.81      1.00      0.90        26

    accuracy                           0.93        80
   macro avg       0.91      0.94      0.92        80
weighted avg       0.94      0.93      0.93        80
```

Hence we tested with Logistic regression and Decision Tree Classification where in the accuracy of Decision Tree classification is 93% compared with Logistic Regression.

Metric	Logistic Regression	Decision Tree Classification
Accuracy	0.925	0.93

Other metrics	`accuracy_score(y_test,y_pred)` → `0.925`
	`conf_mat=confusion_matrix(y_test,y_pred)` `conf_mat` → `array([[48, 6], [ 0, 26]], dtype=int64)`
	`print(classification_report(y_test,y_pred))` → precision / recall / f1-score / support: 0 — 1.00 / 0.89 / 0.94 / 54; 1 — 0.81 / 1.00 / 0.90 / 26; accuracy — / — / 0.93 / 80; macro avg — 0.91 / 0.94 / 0.92 / 80; weighted avg — 0.94 / 0.93 / 0.93 / 80
		`# accuracy score, confusion matrix and classification report of decision tree`
		`dtc_acc = accuracy_score(y_test, dtc.predict(X_test))`
		`print(f"Training Accuracy of Decision Tree Classifier is {accuracy_score(y_train, dtc.predict(X_train))}")`
		`print(f"Test Accuracy of Decision Tree Classifier is {dtc_acc} \n")`
		`print(f"Confusion Matrix :- \n{confusion_matrix(y_test, dtc.predict(X_test))}\n")`
		`print(f"Classification Report :- \n {classification_report(y_test, dtc.predict(X_test))}")`
		Training Accuracy of Decision Tree Classifier is 1.0
		Test Accuracy of Decision Tree Classifier is 0.9333333333333333
		Confusion Matrix :-
		[[70 2]
		[6 42]]
		Classification Report :-
		precision / recall / f1-score / support: 0 — 0.92 / 0.97 / 0.95 / 72; 1 — 0.95 / 0.88 / 0.91 / 48; accuracy — / — / 0.93 / 120; macro avg 0.94 / 0.92 / 0.93 / 120; weighted avg 0.93 / 0.93 / 0.93 / 120

The above table shows that Decision Tree Classification gives better results over Logistic Regression.