PROJECT DEVELOPMENT PHASE

MODEL PERFORMANCE TEST

|  |  |
| --- | --- |
| Date | 10 Nov 2022 |
| Team ID | PNT2022TMID31064 |
| Project Name | Early Detection of Chronic Kidney Disease Using Machine Learning |
| Maximum Marks | 10 Marks |

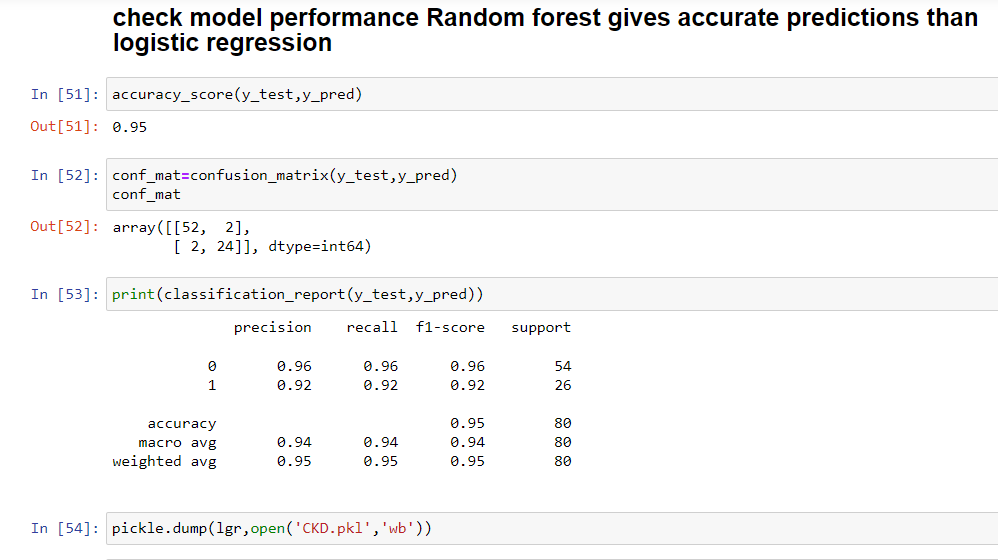
Model Performace Testing :

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: Random Forest Classification



2. Tune the Model

Hyperparameter Tuning:

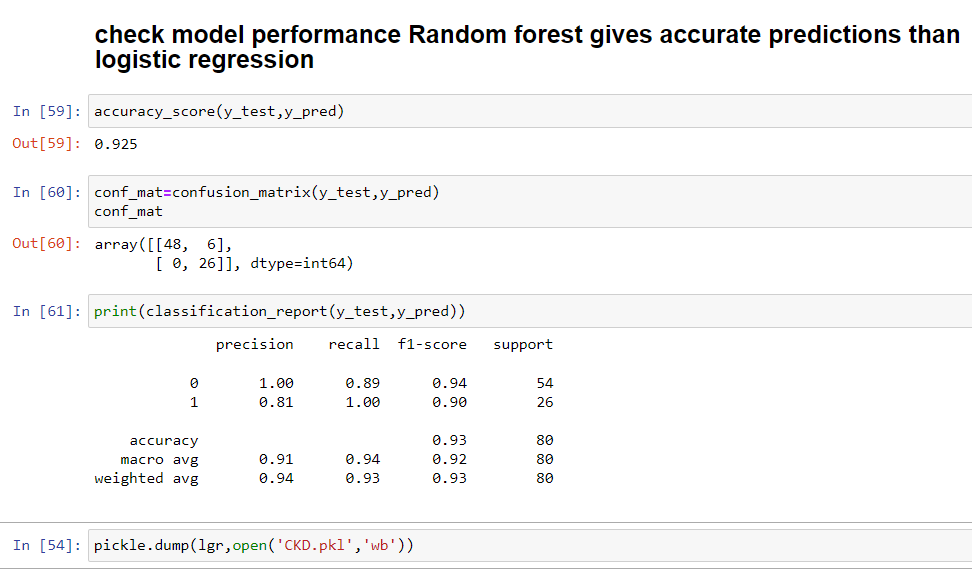
* The number of features is important and should be tuned in random forest classification.
* Initially all parameters in the dataset are taken as independent values to arrive at the dependent decision of Chronic Kidney Disease or No Chronic Kidney Disease.
* But the result was not accurate so used only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



Hence we tested with Logistic regression and Random Forest Classification wherein the accuracy of Random Forest classification is 95% compared with Logistic Regression.

|  |  |  |
| --- | --- | --- |
| Metric | Logistic Regression | Random Forest Classification |
| Accuracy | 0.925 | 0.95 |
| Other metrics |  |  |

The above table shows that Random Forest Classification gives better results over Logistic Regression.