**Model Performance Test**

|  |  |
| --- | --- |
| Date | 19 November 2022 |
| Team ID | PNT2022TMID01939 |
| Project Name | Project – Early Detection of Chronic Kidney  Disease using Machine Learning |
| Maximum Marks | 10 Marks |

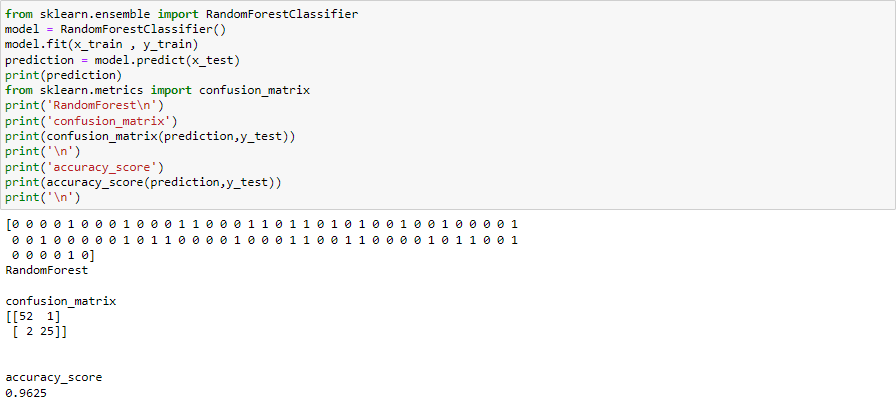
# Model Performance Testing:

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

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No.** | **Parameter** | **Values** | **Screenshot** |
| 1. | Metrics | **Regression Model:**  MAE - , MSE - , RMSE - , R2 score  -  **Classification Model:** Confusion Matrix - , Accuracy Score- & Classification Report  - | See Below |
| 2. | Tune the Model | Hyperparameter Tuning - Validation Method - | See Below |

# Metrics

**Model: Random Forest Classifier**



# Tune the Model

**Hyperparameter Tuning:**

* + The number of features is important and should be tuned in random forestclassification.
  + 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 9 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 | 0.9375 | 0.9625 |
| Other metrics |  |  |

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