**Classification Assignment on Chronic Kidney Disease Prediction**

1. **Identifying Problem Statement:** As for the dataset, it has statistical information and all the necessary input are available to do the prediction for Kidney Disease also the given output is categorical information. So we need to use Machine Learning-Supervised Learning-Classification.
2. **Information about the dataset:** Input columns =24, Output =1, No. of Rows=399

Dataset has both categorical and nominal data. For categorical data conversion is required.

1. **Preprocessing method:** Used Standard Scaler to standardize numerical features and One hot encoding to convert categorical data to nominal data.
2. **Algorithms used:** SVM, Logistic Regression, Decision Tree, Random Forest, KNN, GaussianNB, Gradient Boosting.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.No.** | **Algorithm** | **F1-Score** | **Roc\_auc** | **Github Link** |
| 1 | SVM | 0.99 | 1 | <https://github.com/Thenmozhi6084/ML-Classification-assignment/blob/main/Grid%20SVM-ML%20Classification.ipynb> |
| 2 | Logistic Regression | 0.99 | 1 | <https://github.com/Thenmozhi6084/ML-Classification-assignment/blob/main/Grid%20LogisticRegression-ML%20Classification.ipynb> |
| 3 | Decision Tree | 0.96 | 0.96 | <https://github.com/Thenmozhi6084/ML-Classification-assignment/blob/main/Grid%20DecisionTree-ML%20Classification.ipynb> |
| 4 | Random Forest | 0.99 | 1 | <https://github.com/Thenmozhi6084/ML-Classification-assignment/blob/main/Grid%20RandomForest-ML%20Classification.ipynb> |
| 5 | KNN | 0.96 | 0.99 | <https://github.com/Thenmozhi6084/ML-Classification-assignment/blob/main/Grid%20KNN-ML%20Classification.ipynb> |
| 6 | GaussianNB | 0.98 | 1 | <https://github.com/Thenmozhi6084/ML-Classification-assignment/blob/main/Grid%20GaussianNB-ML%20Classification.ipynb> |
| 7 | Gradient Boosting | 0.99 | 1 | <https://github.com/Thenmozhi6084/ML-Classification-assignment/blob/main/Grid%20GradientBoosting-ML%20Classification.ipynb> |

1. **Final Model:** Since SVM, Logistic Regression, Random Forest gives the best roc\_auc and F1 score, may choose any one from these based on the type of dataset. (linear, non-linear or high-dimensional etc.,)