

1.) Identify your problem statement

The problem statement is to predict the Chronic Kidney Disease based on the several input data's

1. We got the clear requirements and the prediction is based on classification as a inputs[Yes/No]

So We proceeding with Machine Learning

2. We have clear requirement and i/p and o/p data's.

so we choosed on supervised model

3. Based on the given output Classification which is based on Yes/No.

so we have choosen with Classification Models

2.) Tell basic info about the dataset (Total number of rows, columns)

Our dataset has 399 rows × 28 columns

So input-27 columns

output -1 column[Classification => our goal to Predict]

3.) Mention the pre-processing method if you're doing any (like converting string to number – nominal data)

The inputs having Categorical data But Our AI only processed with binary

so we are using One spot Encoding to converting the data's to Binary format.

Also we used standardScaler for better data accuracy after splitting the input and output dataSet.

4.) Develop a good model with good evaluation metric. You can use any machine learning algorithm; you can create many models.

Finally, you have to come up with final model.

After tried with all classification models,

our best model was achieved with 0.99 Accuraccy in SVC and Random Forest Classification.

And F1_score value ->1

5.) All the research values (r2_score of the models) should be documented. (You can make tabulation or screenshot of the results.)

Please check Result_analysis below

6.) Mention your final model, justify why u have chosen the same.

We tried with all hypertuning parameters to achieve The most near by confused matric value.

if it is nearly 1 it is best model. So we tried all classification models and documented then conculed with best model.

Only the best model was saved as **finalized_Classification_model**

Best Model is SVC / Logistic /Rain Forest

Result_analysis

SVC

```
The f1_macro value for best parameter {'C': 10, 'gamma': 'auto', 'kernel': 'sigmoid'}: 0.9924946382275899
Accuracy for SVC Model: 0.9924812030075187
Classification Report for SVC Model:
              precision    recall  f1-score   support

     0           0.98         1.00         0.99         51
     1           1.00         0.99         0.99         82

 accuracy          0.99
 macro avg         0.99         0.99         0.99         133
weighted avg         0.99         0.99         0.99         133

ROC AUC Score for SVC: 1.0
```

Logistic

```
The f1_macro value for best parameter {'penalty': 'l2', 'solver': 'newton-cg'}: 0.9924946382275899
Accuracy for Logistic Model: 0.9924812030075187
Classification Report for Logistic Model:
              precision    recall  f1-score   support

     0           0.98         1.00         0.99         51
     1           1.00         0.99         0.99         82

 accuracy          0.99
 macro avg         0.99         0.99         0.99         133
weighted avg         0.99         0.99         0.99         133

ROC AUC Score for Logistic: 1.0
```

RainForest

```
The f1_macro value for best parameter {'criterion': 'entropy', 'max_features': 'log2', 'n_estimators': 100}: 0.9924946382275899
Accuracy for RF Model: 0.9924812030075187
Classification Report for RF Model:
              precision    recall  f1-score   support

     0           0.98         1.00         0.99         51
     1           1.00         0.99         0.99         82

 accuracy          0.99
 macro avg         0.99         0.99         0.99         133
weighted avg         0.99         0.99         0.99         133

ROC AUC Score for RF: 1.0
```

DecisionTree

```
The f1_macro value for best parameter {'criterion': 'entropy', 'max_features': 'log2', 'splitter': 'best'}: 0.9775556904684072
Accuracy for DC Model: 0.9774436090225563
Classification Report for DC Model:
      precision    recall  f1-score   support

     0       0.98       1.00       0.99        51
     1       1.00       0.99       0.99        82

 accuracy          0.99          0.99          0.99       133
 macro avg         0.99          0.99          0.99       133
 weighted avg      0.99          0.99          0.99       133

ROC AUC Score for DC: 0.9817073170731707
```

KNN

```
The f1_macro value for best parameter {'algorithm': 'auto', 'metric': 'minkowski', 'n_neighbors': 5, 'p': 10}: 0.9478851104269762
Accuracy for KNN Model: 0.9473684210526315
Classification Report for KNN Model:
      precision    recall  f1-score   support

     0       0.88       1.00       0.94        51
     1       1.00       0.91       0.96        82

 accuracy          0.94          0.95          0.95       133
 macro avg         0.94          0.96          0.95       133
 weighted avg      0.95          0.95          0.95       133

ROC AUC Score for KNN: 0.9927068388330942
```