**Assignment 5**

**Chronic Kidney Disease Predictor**

**1.Problem Identification**

1. Domain Selection

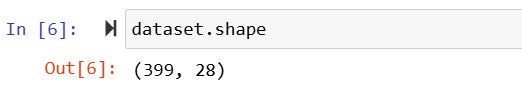
Since the data have lot of numerical rows, I am going with Machine Learning Domain

1. Learning selection

Since the inputs and outputs are clear,its falls under supervised learning.

1. Based on the analysis of the data, the model should predict and classify the patients whether the person will be affected by Chronic Kidney Disease or not . So it is comes under Classification.

**2.Total no of rows and columns of the data**



**3.Pre processing Technique**

We have used two pre processing technique.

**4.Finding better model**

Based on the result of confusion matrix ,classification report,f1\_score, roc\_auc score and accuracy , able to find the model performance. If the above said values should nearest to 1 ,then the model is performing well.

According to the scenario, gradient boosting gives the **best result** as **1.0**

So this is best model.

|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm** | **f1\_score** | **roc\_auc\_score** | **accuracy** |
| Gradient boosting | 1 | 1 | 1 |
| Decision Tree | 0.99 | 0.99 | 0.99 |
| KNN | 0.94 | 1 | 0.94 |
| LogisticRegression | 0.99 | 1 | 0.99 |
| Naive Bayes | 0.97 | 1 | 0.98 |
| Random Forest | 0.99 | 1 | 0.99 |
| SVM | 0.99 | 0.99 | 0.99 |

**5.Deployment**

Since the gradient boosting algorithm gives the best result,I have saved the model for deployment phase.

