**Classification Assignment**

1. Identify your problem statement

Predict whether the patient will get chronic kidney disease or not using the patient information

1. Basic Info about the dataset

Dataset has 25 columns and 399 rows

Input data : 24 column taken as input

Output data : classification(yes or no)

1. Data Preprocessing

Input data: columns rbc,pc(normal/abnormal) - ordinal data to int,

Columns pcc,ba(present/not present) to int

Columns htn,dm,cad,appet,pe,ane (yes/no) tp int

1. Stage 1 : Domain Selection

Machine Learning – Input and output are numbers

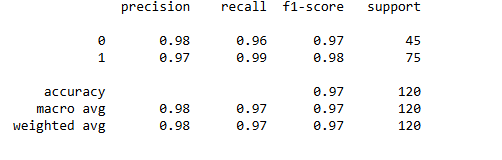
Stage 2 : Learning Selection

Supervised Learning – Req is clear, IP OP is given

Stage 3 : Regression or Classification

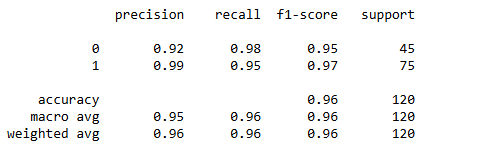
Classification -op is Categorical

1. Algorithms
2. Support Vector Machine



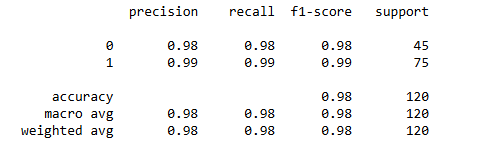
Best Parameter : 

1. Decision Tree

****

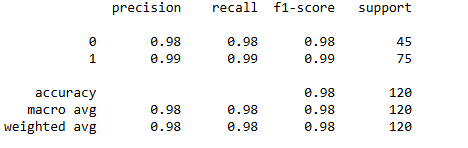
Best parameter - 

1. Random Forest



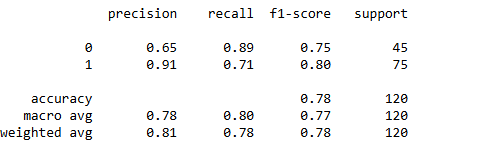
Best Parameter - 

1. Linear Regression



Best Parameter - 

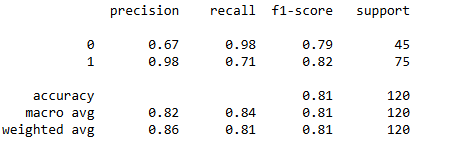
1. K Nearest Neighbor



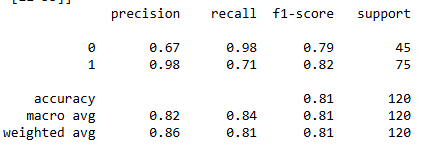
Best Parameter - 

1. Naïve baye’s

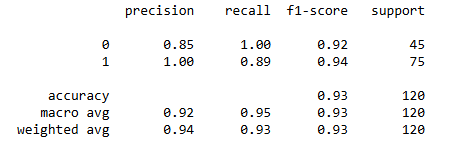
ComplementNB –



MultinomialNB-



BernoulliNB-



The Best Model can be created using Random Forest Classification Algorithm because the accuracy is 98 and weighted avg is 98