Classification Assignment

Problem Statement or Requirement: A requirement from the Hospital, Management asked us to create a predictive model which will predict the Chronic Kidney Disease (CKD) based on the several parameters. The Client has provided the dataset of the same.

1.)Identify your problem statement

In this dataset we have to predict the kidney disease of the client, by their given parameter in dataset

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

Total no of rows : 400

Total on of columns : 26

The columns contains age,
bp,sg,al,su,rbc,pc,pcc,ba,bgr,bu,sc,sod,pot,hrmo,pcv,wc,rc,htn.cad.appet,pe,ane

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

The pre processing method we used is nomila-one heart encoding method

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

In machine learning we used supervised learning and classification model

The best model is RANDOM FOREST

Overall accuracy is : 98

```
print(cm)
```

```
[[50  1]
 [ 1 81]]
```

```
from sklearn.metrics import classification_report
clf_report = classification_report(y_test, y_pred)
```

```
print(clf_report)
```

```
              precision    recall  f1-score   support

           0       0.98      0.98      0.98        51
           1       0.99      0.99      0.99        82

    accuracy                           0.98       133
   macro avg       0.98      0.98      0.98       133
weighted avg       0.98      0.98      0.98       133
```