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
2.) Tell basic info about the dataset (Total number of rows, columns)
3.) Mention the pre-processing method if you're doing any (like converting string to number – nominal data)
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. 5
5.) All the research values of each algorithm should be documented. (You can make tabulation or screenshot of the results.)
6.) Mention your final model, justify why u have chosen the same.

Solution:

1. The client wants a machine learning model to predict Chronic Kidney Disease (CKD) based on input parameters. This is a **supervised learning** problem, specifically a **Classification** problem.

2. The dataset has 399 rows and 25 columns. The provides dataset columns are age, bp, sg, al, su, rbc, pc, pcc, ba, bgr, bu, sc, sod, pot, hrmo, pcv, wc, rc, htn, dm, cad, appet, pe, ane, classification

3. The Categorical columns are sg, rbc, pc, pcc, ba, htn, dm, cad, appet, pe, ane ,classification . LabelEncoder is applied to convert categorical data to numerical data.

4. The Developed models are SVM ,Decision Tree and Random Forest

5.

Model Name	**F1 score**
SVM	0.97
Decision Tree	0.98
Random Forest Regression	1

6. The **Random Forest model** achieved the highest F1 score of **1.** I have chosen it beacause it performs well compared to othermodel. Also it avoids overfitting by aggregating multiple tree.