# Classification Assignment

# Chronic Kidney Disease prediction

## 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

Domain: Machine Learning

Learning Type: Supervised Learning

Algorithm: Classification

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

The dataset used is the Chronic Kidney Disease dataset, which contains 399 rows and 25 columns.

This dataset will be used to create an effective model.

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

Converting categorical data into numerical values.

Using StandardScaler to improve accuracy.

## 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.

In this project, we'll explore six machine learning algorithms to build a predictive model:

- Logistic Regression
- Support Vector Machine (SVM)
- Decision Tree
- Naive Bayes
- K-Nearest Neighbors (KNN)
- Gradient Boosting

## Finalized model

→Gradient Boosting

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

| **ALGORITHM** | **ACCURACY** |
|---|---|
| LogisticRegression | 0.99 |
| SVM | 0.99 |
| KNN | 0.95 |
| DecisionTree | 0.97 |
| Naïve biase | 0.97 |
| Gradientboosting | 1.00 |

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

Gradient Boosting achieved the highest accuracy of 1.00 on this dataset compared to other algorithms,

hyperparameters: learning_rate=0.2, max_depth=3, min_sample_split=2, n_estimators=150, subsamples=1.0.