**Dataset Name:** Chronic Kidney Disease

**Abstract**: This dataset can be used to predict chronic kidney disease and it has been collected at a hospital for a period of nearly 2 months.

**Data Set Characteristics:**

Multivariate

**Number of Instances:**

400

**Area:**

N/A

**Attribute Characteristics:**

Real

**Number of Attributes:**

25

**Date Donated**

2015-07-03

**Associated Tasks:**

Classification

**Missing Values?**

Yes

**Source:**

Source:   
Dr.P.Soundarapandian.M.D.,D.M   
(Senior Consultant Nephrologist),   
Apollo Hospitals,   
Managiri,   
Madurai Main Road,   
Karaikudi,   
Tamilnadu,   
India.

**Problem Statement**

Analyze the data set, investigate and evaluate the result and predict the overall performance.

The dataset has been uploaded to canvas.

The students need to

• Select a method for performing the analytic task

• Preprocess the data to enhance quality

• Carry out descriptive summarization of data and make observations

• Identify relevant, irrelevant attributes for building model.

• Perform appropriate data transformations with justifications

• Generate new features if needed

• Carry out the chosen analytic task. Show results including intermediate results, as needed

• Evaluate the solutions

• Look for refinement opportunities

Following are some points for you to take note of, while doing the assignment:

• Prepare a Report that will supplement the submitted codebase.

• State all your assumptions clearly

• List all intermediate steps and learnings

• Make the report structured and readable.

**Dataset Name: Chronic Kidney Disease**

**Data Set Information:**

**Attribute Nomenclature**  
age - age   
bp - blood pressure   
sg - specific gravity   
al - albumin   
su - sugar   
rbc - red blood cells   
pc - pus cell   
pcc - pus cell clumps   
ba - bacteria   
bgr - blood glucose random   
bu - blood urea   
sc - serum creatinine   
sod - sodium   
pot - potassium   
hemo - hemoglobin   
pcv - packed cell volume   
wc - white blood cell count   
rc - red blood cell count   
htn - hypertension   
dm - diabetes mellitus   
cad - coronary artery disease   
appet - appetite   
pe - pedal edema   
ane - anemia   
class - class

**Attribute Information:**

You need to use 24 + class = 25 ( 11 numeric ,14 nominal)   
1.Age(numerical)   
age in years   
2.Blood Pressure(numerical)   
bp in mm/Hg   
3.Specific Gravity(nominal)   
sg - (1.005,1.010,1.015,1.020,1.025)   
4.Albumin(nominal)   
al - (0,1,2,3,4,5)   
5.Sugar(nominal)   
su - (0,1,2,3,4,5)   
6.Red Blood Cells(nominal)   
rbc - (normal,abnormal)   
7.Pus Cell (nominal)   
pc - (normal,abnormal)   
8.Pus Cell clumps(nominal)   
pcc - (present,notpresent)   
9.Bacteria(nominal)   
ba - (present,notpresent)   
10.Blood Glucose Random(numerical)   
bgr in mgs/dl   
11.Blood Urea(numerical)   
bu in mgs/dl   
12.Serum Creatinine(numerical)   
sc in mgs/dl   
13.Sodium(numerical)   
sod in mEq/L   
14.Potassium(numerical)   
pot in mEq/L   
15.Hemoglobin(numerical)   
hemo in gms   
16.Packed Cell Volume(numerical)   
17.White Blood Cell Count(numerical)   
wc in cells/cumm   
18.Red Blood Cell Count(numerical)   
rc in millions/cmm   
19.Hypertension(nominal)   
htn - (yes,no)   
20.Diabetes Mellitus(nominal)   
dm - (yes,no)   
21.Coronary Artery Disease(nominal)   
cad - (yes,no)   
22.Appetite(nominal)   
appet - (good,poor)   
23.Pedal Edema(nominal)   
pe - (yes,no)   
24.Anemia(nominal)   
ane - (yes,no)   
25.Class (nominal)   
class - (ckd,notckd)