**. Chronic Renal Disease Risk Predictions**

**2. Objective**

A dangerous, rapid medical issue that has globally affected millions of people, Chronic Kidney Disease (CKD) is the process when the renal capacities decrease and later permanently damage kidney functions. Due to the effects of Heart failure, High blood pressure, and diabetes, CKD is a widespread medical diagnosis that can lead to a severe life-threatening disease such as End-stage- Renal- Disease (ESRD). A terminal illness, ESRD, is the process when the kidney stops functioning completely. There are two ways to sustain life after an ESRD diagnosis: receive dialysis treatment or receive a kidney transplant.

My project objective is to merge two different datasets with similar variable classifications. After combining the datasets, my goal is to use regression machine learning techniques to predict particular risk factors that can help identify the signs of CKD in patients. I foresee that I begin using the linear and logistic regression models.

**3. Type of Problem**

**Problem Type:** Upon investigating my two datasets, I observed that my problem type would pivot both classification and regression due to the dataset’s different kinds of variables, sources, and file types.

**4. Dataset Description**

**Dataset Overview:**  
Provide a detailed description of the dataset used in the course, including specific information about its source, collection method, and criteria used for selecting data.

* **Source of Data:**
* Chronic Kidney Disease dataset.
  + The dataset was sourced from Dr.P.Soundarapandian.M.D., D.M., a Senior Consultant Nephrologist from Apollo Hospitals, a medical center located in  Managiri, Madurai Main Road, Karaikudi, T Tamilnadu, India. Research. The data, titled Early Stage of Indians Chronic Kidney Disease (CKD), was created by L.Jerlin Rubini, a research Scholar from Alagappa University. Dr. P. Eswaran, Assistant Professor in the Department of Computer Science and Engineering at Alagappa University, Karaikudi, Tamilnadu, India, guided Rubini.
* Risk factor Predictions of chronic kidney disease dataset

.

* + Electronic health records (HER) from Bangladeshi patients the dataset was sourced from Enam Medical College, a college located in Savar, Dhaka, Bangladesh.
* **Year of Collection:**
* Chronic Kidney Disease dataset.
  + Data was collected from, June to July of 2015.
* Risk factor Predictions of chronic kidney disease dataset
  + Data was collected between 2019 and 2020.
* **Data Collection Method:**

Chronic Kidney Disease Dataset.

* + The data was collected Using electronic health records (EHR) systems, which were manually reviewed by research scholar L.Jerlin Rubini.

Risk factor Predictions of chronic kidney disease dataset

* + The Data was collected through electronic health records (EHR) systems and manually reviewed by data scientists Md. Ashiqul Islam from Diu Journal Analytica R & D Lab and Shamima Akter from the Dept. of Bioinformatics and Computational Biology for accuracy.
* **Purpose of Data Collection:**
* Chronic Kidney Disease Dataset.
  + The data was originally collected to predict the early stage of the chronic kidney disease in Indians.

Risk factor Predictions of chronic kidney disease dataset

* + The data was originally collected to to predict risk factors of the chronic kidney disease in the rural area population of Bangladesh.
  + ."
* **Inclusion Criteria:**
* Outline the specific criteria that determined which data points were included in the dataset.
  + Chronic kidney Disease Dataset:
    - Patients aged 18-80 years old
    - Patients admitted with a diagnosis of high blood pressure, diabetes and anemia.
    - Patients with complete or partial records of blood tests, medical history.
  + Risk factor Predictions of chronic kidney disease dataset:
    - Patients aged 11-74 years old.
    - Patients admitted with a diagnosis of hypertension or diabetes.
    - Patients with complete records of blood tests, stress test, nutrition evaluation medical history.
* **Exclusion Criteria:**  
  Identify any criteria that led to the exclusion of certain data points.
  + Chronic Kidney Disease and Risk factor Predictions of chronic kidney disease dataset:
    - The Chronic Kidney disease dataset has missing values which indicates patients with incomplete medical records or missing data."
    - Patients with diagnose with comorbid conditions such as cancer, or other immune disease such as HIV, serve dehydration, heart failure, Lupus nephritis or Henoch-Schönlein purpura (HSP)
    - Patients under 12 years old or over 85 years old.

**5. Total Number of Estimated Features**

**Features Count:**

Chronic Kidney Disease dataset

* 24 features + class = 25 (11 numeric ,14 nominal)

Risk factor Predictions of chronic kidney disease dataset

* 28 features + class =29

**6. Total Number of Instances/Data Points**

**Instances Count:**

Chronic Kidney Disease dataset.

* 400 instances

Risk factor Predictions of chronic kidney disease dataset.

* 200 instances

**7. Sample Table of Features**

**Feature Description:**  
Please provide a sample table that outlines the features, their types, and units. At a minimum, provide the number of Categorical and Numerical Features.

* Both data sets have similar features in feature abbreviations and feature descriptions, so I have merged the information in this table.

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature Name** | **Description** | **Unit** | **Type** |
| Age | Age of the individual | Years | Numerical |
| Blood Pressure | Individual Diastolic reading | bp in mm/Hg | Numerical |
| Blood pressure limit | Risk factor bp limit | Bp in mm/hg | Numerical |
| Specific Gravity | The measure of water concentrated in urine | Sg-1.005,1.010,1.015,1.020,1.025 | Categorical |
| Albumin | Protein levels in blood and urine | al - (0,1,2,3,4,5) | Categorical |
| .Sugar | Measure of glucose level | su - (0,1,2,3,4,5) | Categorical |
| Red Blood Cells | Measure of low EPO erythropoiesis | rbc - (normal, abnormal) | Categorical |
| Pus Cell | Pyuria (abnormally high num (WBCs) in the urine) | pc - (normal, abnormal | Categorical |
| Pus Cell clumps | kidney infections, urinary tract infections (UTIs) | pcc - (present, not present) | Categorical |
| Bacteria | gut bacteria, which can lead to inflammation | ba - (present,notpresent) | Categorical |
| . Blood Glucose Random | fluctuations in their blood glucose levels, both hypoglycemia and hyperglycemia measure | bgr in mgs/dl | Numerical |
| . Blood Urea | Blood urea nitrogen (BUN) levels measurements | Bu- in mgs/dl | Numerical |
| Serum Creatinine | Measure filter creatinine out of the blood | sc in mgs/dl | Numerical |
| Sodium | Measure sodium intake | sod in mEq/L | Numerical |
| Potassium | High potassium in the blood is called hyperkalemia (heart | pot in mEq/L | Numerical |
| Hemoglobin | complete blood count (CBC) test: check anemia | hemo in gms | Numerical |
| Packed Cell Volume | ratio of the volume occupied by the red cells to the volume of whole blood in a sample arterial blood. | pcv | Numerical |
| White Blood Cell Count | A complete blood count (CBC)  measures the number and quality of white blood cells platelets. | Wbcc-in cells/cumm | Numerical |
| Red Blood Cell Count | A complete blood count (CBC) red blood cells, and platelets. | rbcc in millions/cmm | Numerical |
| Hypertension | Ambulatory blood pressure monitoring (ABPM) | htn - (yes, no) | Categorical |
| Diabetes Mellitus | A1C test  A blood test that measures average blood glucose levels over the past three months | dm - (yes, no) | Categorical |
| Coronary Artery Disease | Noninvasive cardiac stress testing | cad - (yes,no) | Categorical |
| Appetite | Diet Assessment | appet - (good,poor) | Categorical |
| Pedal Edema | swelling in the legs | pe - (yes,no) | Categorical |
| Anemia | Symptoms of low filtration of blood | ane - (yes, no) | Categorical |
| affected | Identify if kidney are affected | Affected 0-no, 1- yes | Categorical |
| stage | Level of CKD | Stage – s1, s2, s3, s4, s5 | Categorical |
| grf | Glomerular filtration rate (GFR) is a measurement of kidney function that's used to assess the severity and progression of chronic kidney disease (CKD): | * Stage 1: GFR greater than 90 mL/min/1.73 m² * Stage 2: GFR 60 to 89 mL/min/1.73 m² * Stage 3a: GFR 45 to 59 mL/min/1.73 m² * Stage 3b: GFR 30 to 44 mL/min/1.73 m² * Stage 4: GFR 15 to 29 mL/min/1.73 m² * Stage 5: GFR less than 15 mL/min/1.73 m | Categroical |
| class | identify individual has CKD | class - (ckd, notckd | Categorical |
|  |  |  |  |