Prediction of Chronic Kidney Disease

**Algorithms Used:**

* Random Forest Classifier

**Colab Notebook:** <https://colab.research.google.com/drive/1KI0be8l3QhG_7e5KnPfqdigDxLYHGIf2?usp=sharing>

**Source:**

<https://www.kaggle.com/code/naniruddhan/asthmaprediction-gnb-balanced-smote/notebook>

**Dataset:**

<https://www.kaggle.com/code/csyhuang/predicting-chronic-kidney-disease/input>

**Rows, columns:** (400, 25)

**About Chronic Kidney Disease (CKD) Prediction Using Health Records**

Chronic Kidney Disease (CKD) is a progressive condition where the kidneys gradually lose their ability to filter waste and excess fluids from the blood. It is often caused by underlying health conditions such as diabetes, high blood pressure, or infections. Early detection is crucial as it enables timely intervention to slow progression and manage associated complications. Predicting CKD using health records offers a data-driven approach to identify at-risk individuals and improve healthcare outcomes.

**Problem Statement**

This study leverages the health records of 400 patients, collected over a two-month period, to predict the presence of CKD. The dataset comprises 24 health-related attributes, with 158 patients having complete records and 242 patients having missing values. Using machine learning techniques, the goal is to develop a predictive model that accurately determines CKD status for patients with incomplete data.

**Attributes Considered for Prediction**

The dataset includes a variety of health indicators and demographic information:

* **Demographic & Basic Health Data**:
* Age (float64): Age of the patient.
* BP (Blood Pressure, float64): Average blood pressure over the period.
* **Laboratory Measurements**:
* SG (Specific Gravity, float64): Measure of urine concentration.
* AL (Albumin, float64): Protein levels in urine.
* SU (Sugar, float64): Sugar levels in urine.
* RBC (Red Blood Cell Count, float64): Indicates anemia or kidney function.
* PC (Pus Cell Count, float64): Suggests infection or inflammation.
* BGR (Blood Glucose Random, float64): Random blood sugar levels.
* BU (Blood Urea, float64): Indicator of kidney function.
* SC (Serum Creatinine, float64): Waste product level in blood.
* SOD (Sodium, float64) & POT (Potassium, float64): Electrolyte levels.
* HEMO (Hemoglobin, float64): Indicates oxygen-carrying capacity of blood.
* PCV (Packed Cell Volume, object): Blood test result.
* WC (White Blood Cell Count, object): Infection indicator.
* RC (Red Blood Cell Count, object): Measures anemia severity.
* **Clinical Conditions**:
* HTN (Hypertension, float64): Presence of high blood pressure.
* DM (Diabetes Mellitus, float64): Indicates diabetes status.
* CAD (Coronary Artery Disease, float64): Indicates heart disease.
* Appetite, PE (Pedal Edema), ANE (Anemia): Clinical symptoms affecting CKD.
* **Target**:
* Class (float64): CKD status (1 for CKD, 0 for non-CKD).

**Algorithms Model and Accuracy**

|  |  |
| --- | --- |
| **Models** | **Scores** |
| Random Forest Classifier | 0.8884 |