

Chronic Kidney Disease Analysis

Instructions to RUN THE CODE which is in Modularized Reusable form:

Create a folder and place all the 5 python files (CKD_Analysis.py, import_data.py, Data_Preprocessing.py, PCA_Analysis.py, Classification.py) within that folder. Also place the input file (csv_result-chronic_kidney_disease.csv) in that folder.

For sake of simplicity, use the attached csv file.

Steps:

- 1) Run CKD_Analysis.py where it will call the import_data module and prompt the user (dialog box will open) to select the input file (csv_result-chronic_kidney_disease.csv as mentioned above), but the code is such that any file format will be accepted, and it will be converted into a dataframe.
- 2) Once the file is loaded, Data_Preprocessing module will be called where the data will be processed i.e., few unwanted columns are removed due to lack of data and the '?' are replaced by NaN/0. (other few functions are also there for similar kind of datasets)
- 3) Now, since the data is in correct format, PCA analysis module will be called, and major risk factors are identified.
- 4) Lastly, we call the classification module to identify the potential sub-types of ckd.

Solution:

- a) From the PCA Analysis, we conclude that the risk factors of chronic kidney disease are : **Sugar, Blood Glucose, Blood Urea, Serum Creatinine, Hemoglobin, Packed Cell Volume, Sodium, Potassium**
- b) From the classification model using XGBoost algorithm, we conclude that the potential sub-types of ckd are **Hypertension, Diabetes Mellitus, Pedal Edema, Anemia, with model accuracy 87.5%**