

### 9.4 Exercise do it yourself:

*Download Source Data Set from GitHub link* : <https://github.com/Smartbrain2024/Mastering_AI_2.git>

**Chapters/Chp\_09/9.4/DIY\_Exercise/kidney\_disease.csv**

The data was taken over a 2-month period in India with 25 features ( e.g. , red blood cell count, white blood cell count, etc. .). The target is the 'classification', which is either 'ckd' or 'notckd' - ckd=chronic kidney disease. There are 400 rows.

The data needs cleaning: in that it has NaNs, and the numeric features need to be forced to floats. We were instructed to get rid of ALL ROWS with Nans, with no threshold - meaning, any row that has even one NaN, gets deleted.