**Stroke Prediction Dataset**

**(Data Dictionary)**

Context

According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths.  
This dataset is used to predict whether a patient is likely to get stroke based on the input parameters like gender, age, various diseases, and smoking status. Each row in the data provides relevant information about the patient.

Attribute Information

1) id: unique identifier  
2) gender: "Male", "Female" or "Other"  
3) age: age of the patient  
4) hypertension: 0 if the patient doesn't have hypertension, 1 if the patient has hypertension  
5) heart\_disease: 0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease  
6) ever\_married: "No" or "Yes"  
7) work\_type: "children", "Govt\_jov", "Never\_worked", "Private" or "Self-employed"  
8) Residence\_type: "Rural" or "Urban"  
9) avg\_glucose\_level: average glucose level in blood  
10) bmi: body mass index  
11) smoking\_status: "formerly smoked", "never smoked", "smokes" or "Unknown"\*  
12) stroke: 1 if the patient had a stroke or 0 if not  
\*Note: "Unknown" in smoking\_status means that the information is unavailable for this patient

Data Changes / Manipulations

1. Data in ‘healthcare-dataset-stroke-data’ is in its original form and should only be changed in the Jupyter Notebook
2. Data in ‘healthcare-dataset-stroke-data(cleaned)’ has been cleaned at the csv level and will NOT need to go through the data cleaning steps in Jupyter Notebook

Changes made to ‘healthcare-dataset-stroke-data(cleaned)’ are:

* Column ‘gender’
  + Male = 0
  + Female = 1
  + Other = 2
* Column ‘ever\_married’
  + Yes = 1
  + No = 0
* Column ‘work\_type’
  + Private = 3
  + Self-employed = 4
  + Govt\_job = 2
  + Children = 1
  + Never\_worked = 0
* Column ‘Residence\_type’
  + Urban = 2
  + Rural = 1
* Column ‘smoking\_status’
  + formerly smoked = 1
  + never smoked = 2
  + smokes = 3
  + Unknown = 0

Replaced N/A values in the ‘bmi’ column with the mean value of the ‘bmi’ column.

Acknowledgements

(Confidential Source) - *Use only for educational purposes*  
If you use this dataset in your research, please credit the author.