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 relavant information about the patient.

Attribute Information

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

Acknowledgements

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

<https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>