**Documentation Of Dataset**

**1. Size of the Dataset:** Number of columns 11 and number of rows is 250.

**2. Description of features with their units**: Brain stroke is mainly based on some terms. Here I used this terms as features. In this dataset some features are used to predict the stroke. They are gender, age, hypertension, heart disease, marital status, work type, residence type, average glucose level, BMI(Body Mass Index), smoking status and the target class which is ‘Stroke’. Here are 250 rows in this dataset. Here gender class is labelled as ‘Male’ and ‘Female’. If it is male it will be shown 1 otherwise it will be shown 0. This part is handled in code. Same for the class marital status, work type, residence type, smoking status.

**3. Citations of the websites:** [Stroke Prediction Dataset | Kaggle](https://www.kaggle.com/fedesoriano/stroke-prediction-dataset)

**4. Distribution of the labels:** In target class there are 106 times labelled 1 which means 106 persons have stroke and 144 times labelled 0 which means 144 persons don’t have stroke among 250 rows of data.