Stroke Prediction

Lin Yang ly2565

**Introduction**

Stroke is a medical emergency and a brain attack that interrupts blood supply and oxygen to the brain. According to the National Stroke Association, stroke is the fourth leading cause of death in the US, and a leading cause of serious and long-term adult disability. The high stroke mortality has caused significant cost burden, including healthcare services and medications, thus it is important to study and identify risk factors for stroke. There are many risk factors for stroke, such as age, sex, race, hypertension, smoking status, diabetes, etc. The dataset used in this study is collected to predict the probability that a patient gets stroke based on predictors like gender, age, hypertension, work type, and BMI. Each row in the dataset provides relevant information about the patient.

All columns we have are shown below:

* id: unique identifier
* gender: "Male", "Female" or "Other"
* age: age of the patient
* hypertension: 0 if the patient doesn't have hypertension, 1 if the patient has hypertension
* heart\_disease: 0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease
* ever\_married: "No" or "Yes"
* work\_type: "children", "Govt\_jov", "Never\_worked", "Private" or "Self-employed"
* Residence\_type: "Rural" or "Urban"
* avg\_glucose\_level: average glucose level in blood
* bmi: body mass index
* smoking\_status: "formerly smoked", "never smoked", "smokes" or "Unknown"\*
* 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