

DATASET

BASIC INFO:

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.

ABOUT THE PROBLEM

Our top priority in this health problem is to identify patients with a stroke.

ATTRIBUTES

- 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

EXTRA DATA

NUMBER OF SAMPLES:

1 which has been divided into training and test data

CLASSES:

Contains

1. Binary class: only two classes e.g. (labels- yes/no)
2. Multi class: more than two classes e.g. (dress, bag, footwear, jewellery)

DISTRIBUTION OF CLASSES

1. Binary Class: gender, hypertension, heart_disease, ever_married, Residence_type, stroke
2. Multi Class: age, id, work_type, avg_glucose_level_mbi, smoking_status

CLEANING THE DATASET

```
In [1]: import csv
with open('healthcare-dataset-stroke-data.csv', 'r') as inp, open('dataset.csv', 'w') as out:
    writer = csv.writer(out)
    for row in csv.reader(inp):
        if (row[9] != 'N/A' and row[10] != 'Unknown'):
            writer.writerow(row)
```

NOTE: data.csv is the dataset obtained after cleaning