Stroke Prediction Dataset

[Selene Hakobyan]

Problem Statement

Is it possible to predict whether or not an individual will suffer a stroke based on input parameters such as gender, age, various diseases, smoking status, etc.? What factors affect stroke occurrence and can we predict the likelihood of a stroke?

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 a 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.

Criteria for Success

Criteria for success for this project would be the correctness of predicting whether or not an individual had a stroke and being able to accurately predict the likelihood of a stroke based on input parameters such as age, gender, etc.

Scope of solution space

The main focus is on identifying which factors make it more likely for individuals to suffer from a stroke. The possible factors include the following: gender, age, bmi, smoking status, work type, etc.

Constraints

The dataset might be imbalanced; there is a small percentage of individuals who had a stroke.

Stakeholders

- Medical professionals who can use this data to identify individuals who are more likely to suffer from a stroke and take action steps to prevent its occurrence.
- Patients
- Caregivers
- Hospitals

Data Sources

Data from Kaggle: Stroke Prediction Dataset