The dataset was obtained through the Kaggle.com website by user FEDESORIANO in 2020. The owner authorised this dataset for educational purposes.

This dataset contains detailed records of 5110 patients in an anonymous hospital. As mentioned in the proposal, some of these records will be removed as they do not contain enough information for the learning model.

It is impossible to fully verify its accuracy as its source and the method of information collection is confidential although the author did confirm that it was collected form a hospital database. In addition, 5110 might not be enough to represent the population so the accuracy and representativeness of the learning model will suffer

Variable description:

* ID: Uniquely identify a patient
* gender: Subtype of population the patient is identified with
* age: How old the patient is
* hypertension: Whether the patient suffered from high blood pressure
* heart\_disease: Whether the patient suffered from heart diseases
* ever\_married: Whether the patient have married
* work\_type: Patient career
* residence\_type: Where the patient live
* avg\_glucose\_level: average glucose level in blood
* bmi: Body mass index
* smoking\_status: The patient’s experience with smoking
* stroke: Whether the patient suffers from strokes

pca, svm look into

stroke

#https://www.kaggle.com/datasets/prosperchuks/health-dataset

2 things

1 normal

Age and sex

2 lifestyles

Alcohol

Diet

Exercise

Income

Marital

Sleep

Health coverage

Mental

smoking