

Stroke Prediction

What is the framing question of your analysis, or the purpose of the model/system you plan to build?

Using the dataset to predict stroke
model type is classification

Who benefits from exploring this question or building this model/system?
Patients and doctors

Data Description:

- What dataset(s) do you plan to use, and how will you obtain the data?

<https://www.kaggle.com/fedesoriano/stroke-prediction-dataset/version/1>

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

Number of columns 12

Number of rows 5110

- What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?

Each row is patient

With id, gender, age, hypertension, heart_disease, ever_married, work_type, Residence_type, avg_glucose_level, bmi, smoking_status, stroke features.

Features:

- 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

- If modeling, what will you predict as your target?

Target: stroke

Predict whether he will have a stroke or not.

Tools:

- How do you intend to meet the tools requirement of the project?
use the tools [Numpy](#), [Panda](#), [Scikit-Learn](#), [Matplotlib](#).
- Are you planning in advance to need or use additional tools beyond those required?
[Seaborn](#)