

Project Proposal

Stroke Prediction

NOJOURD ABDULLAH ALSHEDOKHI

Question/need:

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

The purpose of the model is to predict whether a patient is likely to get stroke based on the input parameters.

- **Who benefits from exploring this question or building this model/system?**

HealthCare sector, Hospitals, Doctors, and patients.

Data Description:

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

I download the dataset from Kaggle website. It is containing 5110 observations with 12 attributes.

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

This dataset contains an individual attribute about patients, and group unite when comparing two groups, those that had a stroke and those who did not.

1. The names and feature labels:

- **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

2. Categorical and Numerical features are present

- Categorical Features: gender, ever_married, work_type, Residence_type, smoking_status
- Binary Numerical Features: hypertension, heart_disease, stroke
- Continuous Numerical Features: age, avg_glucose_level, bmi
-

3. bmi feature has missing values

• **If modelling, what will you predict as your target?**

Supervised learning (classification model)

Tools:

• **How do you intend to meet the tools requirement of the project?**

Data Processing: Pandas, Numpy, Scipy

Modelling: Scikit_Learn

Visualization: Seaborn, Matplotlib.

• **Are you planning in advance to need or use additional tools beyond those required?**

Yes

MVP Goal:

What would a minimum viable product (MVP) look like for this project?

Does this patient will have a Stroke?

Does married people gets more stroke than unmarried people?

Does age have direct impact on stroke?