

Project Proposal

This repo is T5 Bootcamp requirement.

What kind of people who are likely to have stroke?

Based on some research and studies, stroke is considered as one of the leading causes of death worldwide. This project is used to predict patients with high risk to get stroke depending on certain attributes.

The project objective is predicting whether a patient will have stroke based on some given features.

Dataset

To accomplish the purpose of project, the “**Stroke Prediction Dataset**” will be used which can be found at [Kaggle](#).

This dataset has the following attributes:

- ID
- Gender
- Age
- Hypertension
- Ever Married
- Work Type
- Residence Type
- Avg Glucose Level
- BMI
- Smoking Status
- Stroke

Here is a data sample shown in the following table:

id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
9046	Male	67	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1

The most important features for this project will be identified after data modeling and evaluation.

Tools:

Some tools will be used such as: sklearn, pandas, numpy, matplotlib

To Do:

Data will be explored with EDA phases. Then a model will be used to fit the data. NOTE: some features might be increased or changed and the model as well.