Stroke Prediction

Abstraction:

A stroke occurs when the blood supply to part of your brain is interrupted or reduced, preventing brain tissue from getting oxygen and nutrients. Brain cells begin to die in minutes. **The purpose of this project** is to use classification models to predict whether a given person had a stroke or not, and find the main risk factor for stroke to create a prediction system to predict the stroke in its early stages.

Question/need:

The main question of this project is "**what is the main risk factor for stroke**?"

Additional questions:

- Does high blood pressure have an impact on stroke?
- Do people work in private sector has more chance of stroke?
- Does married people gets more stroke than unmarried people?
- Does age have direct an impact on stroke?
- Dose smoking have direct an impact on stroke?
- Do people work in private sector has more chance of stroke?

Data Description:

I will work with data provided by Kaggle, the dataset contains **5110 recorders** with 12 features for each, 7 of which are categorical. 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.

Tools:

- NumPy and Pandas for data manipulation
- Scikit-learn for modeling
- Matplotlib and Seaborn for plotting

Model:

I will select the appropriate model for this dataset from these classification models (Logistic Regression, Decision Tree, K Nearest Neighbour, and Random Forest Classifier) based on their performance, so I will tune the hyperparameters with the help of GridSearch to get the model.

MVP Goal:

The MVP will achieve two goals of this project, which are:

1- Identify patients with a stroke:

Achieve it by using the suitable model which has the highest accuracy to predict the patient had a stroke or not.

2- What the factor has the most impact on stroke?

Achieve it by finding the correlation between features and the target then visualize them using Seaborn and Plotly libraries.