**Team Sage**

**Topic Proposal**

The Brain Stroke data set is from Kaggle. It provides information on how various factors affect a person causing a stroke. Presently there are a total of 4,981 entries in the dataset with 10 features which include: gender, age, hypertension, heart disease, marital status, type of work, residence type, average glucose level, body mass index, smoking status, and stroke. Among these, 5 features are categorical, and 6 are numeric values. The research topic for this project is: What are the most accurate predictors of the likelihood of having a brain stroke?

**SMART Questions**

* What are the most accurate predictors of the likelihood of having a brain stroke?
* What are the main factors that can induce a stroke?
* What combination of factors has a higher chance of including a stroke in an individual?
* Do all the factors provided help us understand how a stroke can occur?
* How does the length of a symptom influence the probability that a brain stroke will occur?
* Can we make predictions for any other health issues using similar data?

**Source**

**The source of our data set**: Brain Stroke Prediction on Kaggle:

**(**<https://www.kaggle.com/datasets/zzettrkalpakbal/full-filled-brain-stroke-dataset?datasetId=2343381&sortBy=voteCount>**)**

**Number of observations**: 4,891

**GitHub**

The link to our GitHub repository: <https://github.com/sowmyamaddali/DATS6101_Project_Team_Sage>