Project Title: **Brain Stroke Prediction Dataset**

Name: Bharath Posa
ID: 11594402

Name: Gagan Sai Ram Anvesh Achanta
ID: 11447940

Name: Vignesh Pasula
ID : 11603369

Name : Nithish Reddy Boddhi Reddy
ID: 11555383

Idea description
The idea behind this project is to come up with a machine learning model the can predict the Chances of getting Brain Stroke provided the inputs of various health parameters.

Goals and Objectives:

i.   Visualize the various parameters and do analysis on each parameter to see the significance of failure
ii.   To do the exploratory data analysis on the data to understand the parameters and their significance
iii.   To apply statistical procedures to build insights on the dataset.
iv.   To perform P test analysis on the data
v.   To come up with machine learning model in order to predict for new inputs.

Motivation
According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths.
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. Each row in the data provides relevant information about the patient.

Significance
We can use this model to predict whether a person will have a brain stroke or not. It also generates observations on the data that can be useful in understanding the key parameters

(vitals) that need to be checked to predict the chances of having a brain stroke, and if the chances are high, the patient can take precautionary measures to avoid having one.

Literature Survey

Brain Strokes are leading cause of death. There has been various studies on predicting it . In order to classify stroke various ml models that are used to build classifier were reviewed . We looked at various research works to understand how the data has been interpreted. The information from Mayo forms basis for understanding the health issue. The WHO organization data provides insight into demographics of the disease at wide scale. The ucla information was useful to understand the dynamics of the stroke. Links are provided for the resources in reference section.

Dataset:
https://drive.google.com/drive/folders/1AHPkgPG8xkbMCdIf30lB7vmmlYPorqwa?usp=sharing

Features
    List of features in the given dataset are as follows

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

Expected outcome

To come up with a machine learning model to predict patient likeliness to have a brain stroke given his vitals information.

References

https://www.mayoclinic.org/diseases-conditions/stroke/symptoms-causes/syc-20350113
https://www.cdc.gov/stroke/about.htm
https://www.uclahealth.org/medical-services/radiology/clinical-services/stroke-brain-attack
https://www.who.int/southeastasia/news/detail/29-10-2016-prevent-brain-stroke
https://www.who.int/health-topics/brain-health#tab=tab_1