Stroke Prediction

# **Vinay Nagaraj DSC680 - T302 Applied Data Science (2217-1)** <https://vinaynagaraj88.github.io/DataScience_Portfolio>

# Which Domain?

For my third project, I am planning to work in the Healthcare domain. I have not worked on a project in healthcare domain this entire course and I felt this will be a good topic to research and learn about.

According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths. A stroke happens when blood stops flowing to any part of your brain, damaging brain cells. The effects of a stroke depend on the part of the brain that was damaged and the amount of damage done.

A Report from the American Heart Association informs that an average, someone in the US has a stroke every 40 seconds. There are about 795,000 new or recurrent strokes each year, based on 1999 data. On average, someone dies of a stoke every 3 minutes and 33 seconds in the US. There are about 405 deaths from stroke each day, based on 2018 data.

In general, stroke is more likely to occur in the elderly, and stroke can lead to cerebral dysfunction such as hemiplegia, mispronunciation, and lack of consciousness. Properly managing and treating adjustable risk factors such as hypertension, smoking, diabetes, and obesity can decrease stroke occurrences. Stroke is a treatable disease, and if detected or predicted early, its severity can be greatly reduced.

Through this project, I intend to consider all the relevant information about the patient such as gender, age, various diseases, and smoking and build predictive analytics techniques that would predict the patients with high risk and is likely to get stroke. This helps in providing the advanced warning to alert the patients so that they can apply proper precautions and possibly the prevent the stroke.

Below are several references I am planning to refer as part this project:

1. <https://www.health.harvard.edu/heart-health/stroke-after-a-heart-attack-whats-the-risk>
2. <https://www.healthline.com/health/stroke-vs-heart-attack>
3. <https://www.ahajournals.org/doi/10.1161/STROKEAHA.120.031295>
4. <https://www.sciencedirect.com/science/article/abs/pii/S0828282X13003607>
5. <https://medium.com/geekculture/stroke-prediction-d26c15f9d1>
6. <https://www.nature.com/articles/s41598-021-89434-7>
7. <https://www.dicardiology.com/article/how-artificial-intelligence-can-predict-and-detect-stroke>
8. <https://www.datasciencecentral.com/profiles/blogs/stroke-prediction-using-data-analytics-and-machine-learning>
9. <https://www.techexplorist.com/improving-stroke-recovery-prediction-using-machine-learning/40015/>
10. <https://avicenna.ai/stroke-detection-ai-tool-approved-for-medicare-ntap/>

# Which Data?

I am planning to use the below dataset from Kaggle which contains details about patients with features such as gender, age, various diseases, and smoking status. The target variable in this dataset is a binary variable reflecting the fact whether the patient had a stroke or not. “stroke”: 1 if the patient had a stroke or 0 if not. I am planning to use this data to train the model as part of this project.

Dataset Link - <https://www.kaggle.com/fedesoriano/stroke-prediction-dataset>

The dataset contains 5110 observations with 12 attributes.

# Research Questions? Benefits? Why analyze these data?

Below are some of the research questions I would like to find answers for which would provide a lot of sights about the causes of Stroke and how to predict and prevent.

* Which are the most important features which helps us to predict a Stroke?
* Does age has impact on strokes? and How is this parameter distributed?
* Does body mass index and glucose levels in a person, propel a heart stroke?
* Smoking can induce Stroke, is it true?
* Heart with a Heart Disease is prone to Stroke, is it true?
* Workload results in high blood pressure and that could lead to Stroke, is it true?
* Males are most susceptible to strokes due to high work related stress, is it true?
* Which is the best prediction model to predict Stroke?
* What is the accuracy of the available classification models in predicting Stroke?

# What Method?

* I am planning on using Python language and work on Jupyter notebook for this project.
* I will initially clean the dataset for any missing values or outliers.
* Perform some EDA on the dataset to understand the trend of the data.
* I will perform Feature reduction/selection and then build or use some of the classification models to make predictions on employee attrition.

# Potential Issues?

My initial review of the data and the columns do provide some good insights and I don’t see any major issues or roadblocks at this point.

# Concluding Remarks

If stroke can be predicted at an early stage there is 4% lower risk of in-hospital death, 4% better odds of walking independently after leaving the hospital and also 3% better odds of being sent home instead of to an institution.

The aim of this project is to analyze data related to the patients and gain insight of the reasons for Stroke and its prediction. It is important to focus on patients identified by the prediction model. All of us can greatly benefit if a stroke can be predicted at an early stage and proper medication and steps are taken to cure it.