

Data Science Career Track

Capstone Two: Project Proposal

Stroke Prediction Dataset

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

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 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.

Proposal:

Using a dataset of patient attributes to determine the likelihood of stroke. Preventative measures could be taken to save patients from the detrimental side effects of having a stroke.

Problem Identification:

Problem Statement formation

What patterns in a patient's attributes make them more or less likely to have a stroke?

How can we track that and lower overall healthcare costs to the patient, and what measures could be taken to prevent a patient from getting a stroke?

Context

Stroke is the 2nd leading cause of death globally, 11% of total deaths. We would like to be able to predict whether or not a patient is likely to get a stroke based on a patient's health attributes.

Criteria for success

The key criteria that will make this successful are the input parameters like gender, age, various diseases, and smoking status; and how we can determine if a patient might be likely to get a stroke.

Scope of Solution Space

What is focus is understanding how each of the parameters might influence a patient's

likelihood to have a stroke.

What will we focus on specifically in the 12th column focuses on if the patient had a stroke or not, we can see what parameters that patient had and build a model off of it.

Constraints within solution space:

The constraints that we may face is the dataset could be relatively small compared to the number of people who actually die from stroke related causes. We also may need to understand more about relative health compared to an unhealthy person, ex: understanding a good BMI verses a bad BMI patient.

Stakeholders:

Public Health, Healthcare analytics

Key Data Sources:

Username: fedesoriano on Kaggle

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