

Classification Project Proposal

Josh Jingtian Wang

4/2/2022

Question/Need:

A stroke is a medical condition that cuts off the oxygen supply to the brain. In 2015, stroke was the second most frequent cause of death after coronary artery disease. Due to its deadly and sudden nature, it is important to be able to predict the onset of strokes. Here I propose a classification model to predict the onset of strokes using features that can easily be collected from patients.

Data Description:

Date from: <https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>

Input variables include physical metrics of patients such as age, BMI, health history.

The output variable is whether the patient had a stroke (1 or 0).

Tools:

Data cleaning and prediction will be done using pandas and sklearn in Python.

App will be made with Flask.

MVP Goal:

Building a prediction model using kNN and logistic regression.