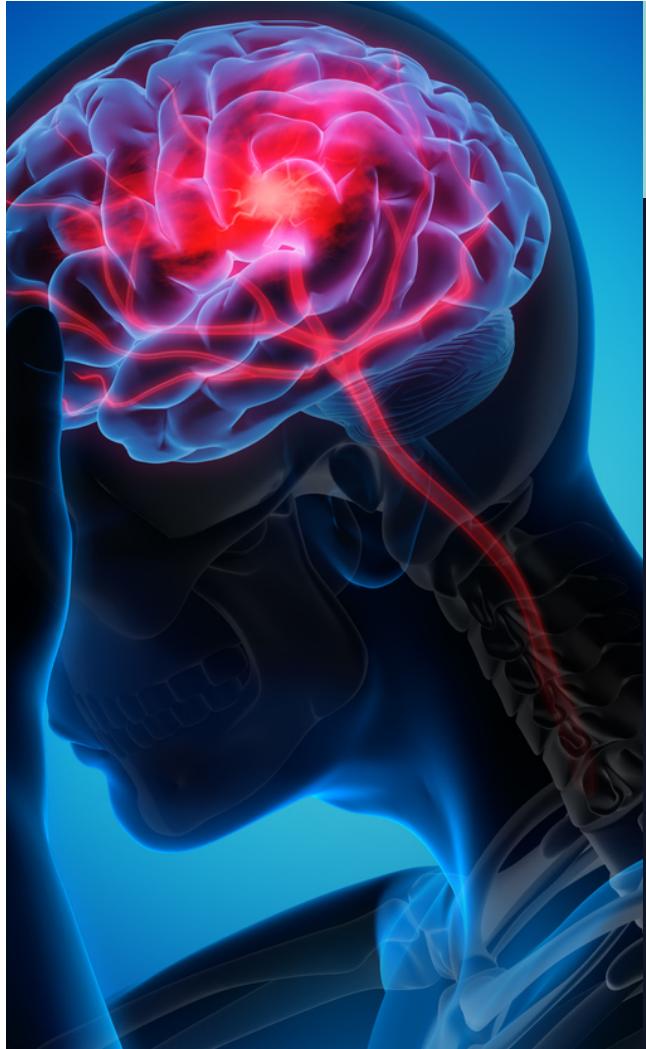


Stroke Classification

Jordana Tepper, Troy Hendrickson, Viktoria Szontagh



What is a “Stroke” ?

“ A stroke, sometimes called a brain attack, occurs when something blocks blood supply to part of the brain or when a blood vessel in the brain bursts. ”

CDC

Two types of Strokes

Ischemic stroke

- WHEN A BLOOD CLOT BLOCKS THE BLOOD FLOW IN THE BRAIN
- MOST STROKES ARE ISCHEMIC

CDC

Hemorrhagic stroke

- WHEN AN ARTERY IN THE BRAIN LEAKS BLOOD OR RUPTURES (BREAKS OPEN).

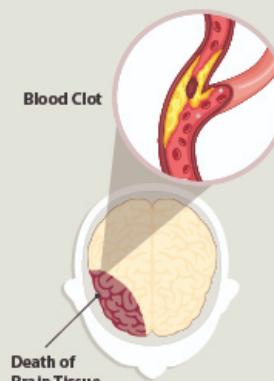
CDC

Stroke

A stroke, sometimes called a brain attack, happens in one of two ways:

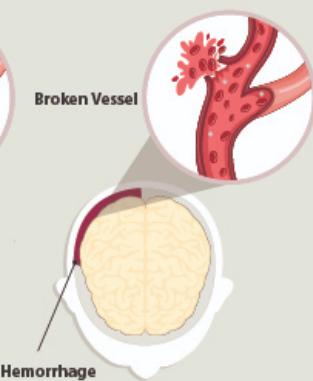
1 Blocked Artery

An ischemic stroke occurs when a blood clot blocks the blood flow in an artery within the brain.



2 Ruptured Artery

A hemorrhagic stroke occurs when a blood vessel bursts within the brain.



Business Understanding

- **STAKEHOLDERS**

Mount Sinai Hospital

- **THE PROBLEM**

According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths.

- **THE PROJECT**

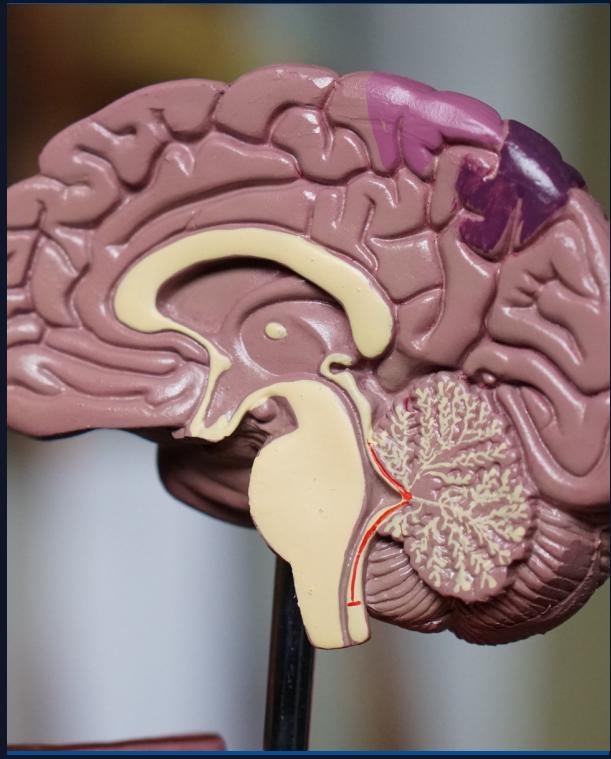
Develop a model that determines whether a person is likely to have a stroke or not have a stroke using available data

- **THE GOAL**

Introduce the model that arbitrates the best results to identify patients in need of extra screening before surgery



The Data



- FEATURES
- MISSING DATA
BMI (Body Mass Index)
- IMBALANCE
96% - No stroke
4% - Stroke

- KAGGLE 'STROKE PREDICTION DATA SET'

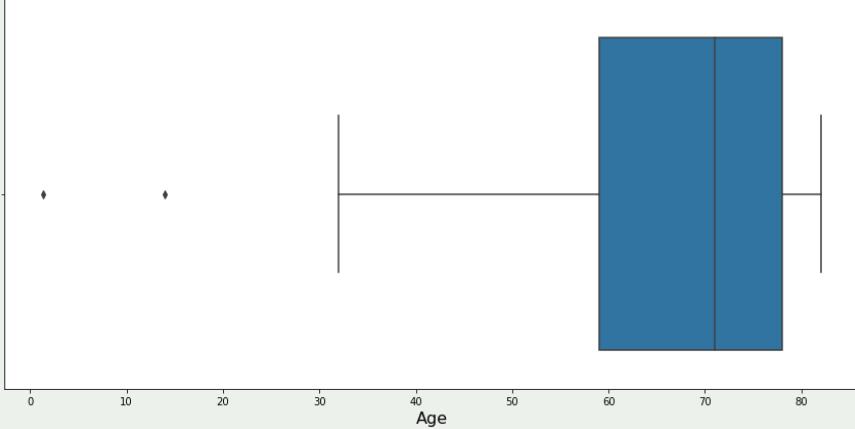
5k+
ROWS

11
COLUMNS

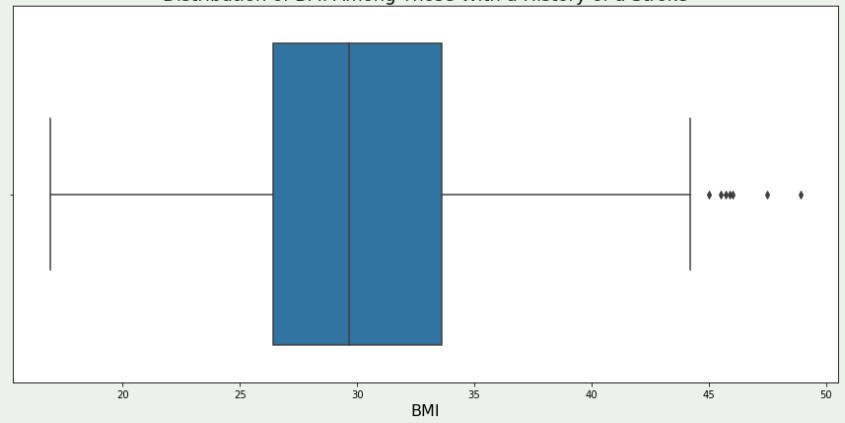
The Metrics That Matter

Data Processing

Distribution of Ages Among Those With a History of a Stroke

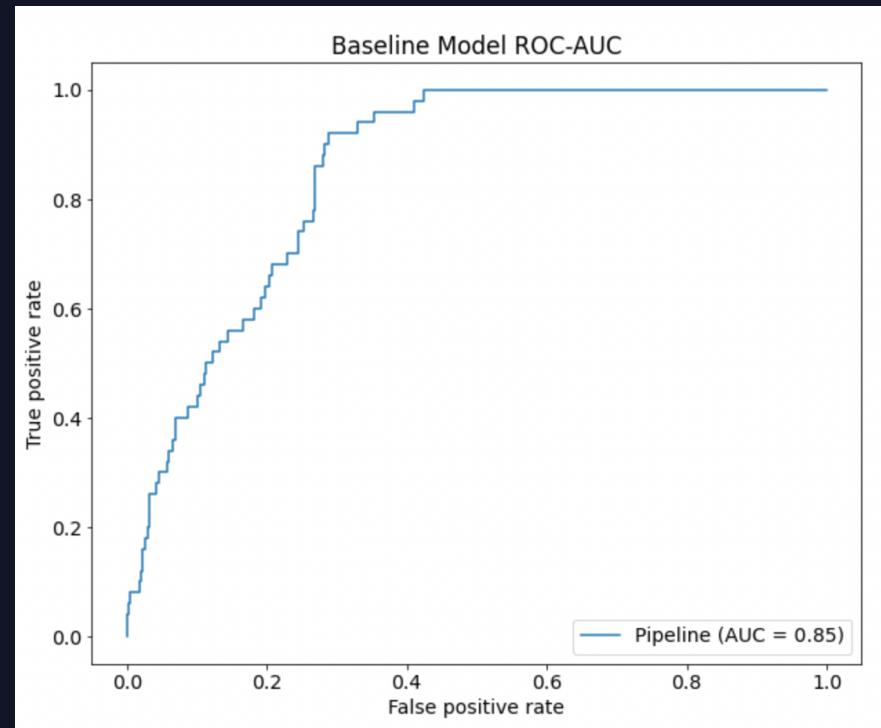
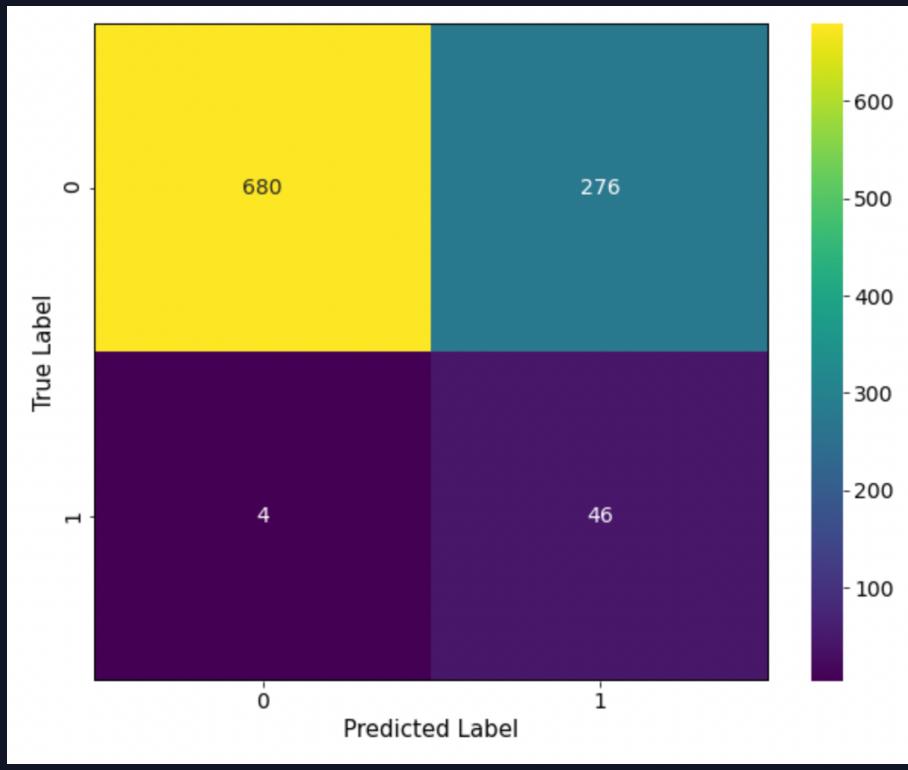


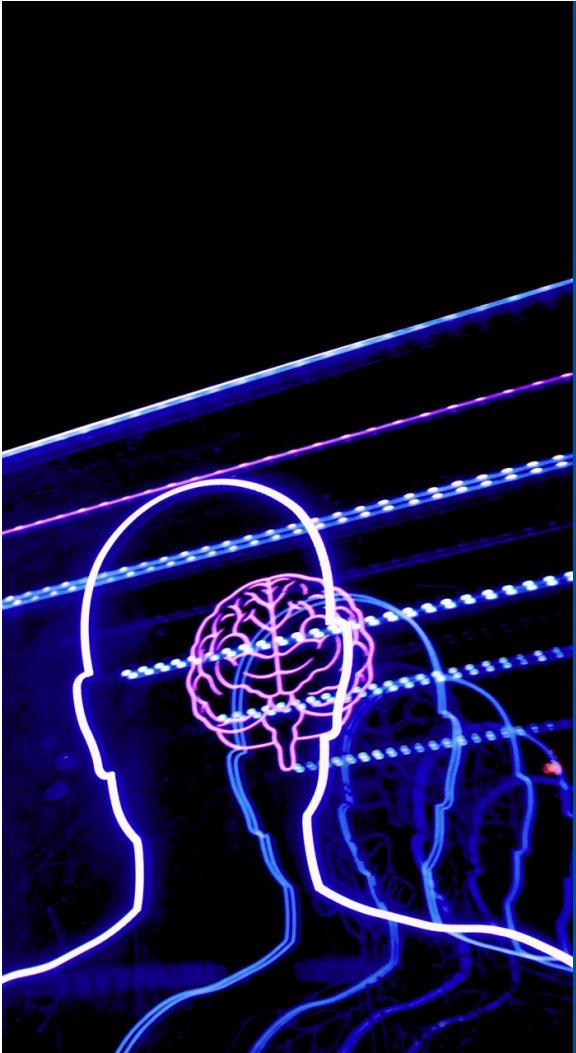
Distribution of BMI Among Those With a History of a Stroke



Best Model

Logistic Regression - Optimal Threshold





✓ CLASSIFICATION MODEL: LOGISTIC REGRESSION

Included optimal threshold for best results

✓ RECALL SCORE: 92 %

The main metric used to determine the accuracy of our model
A false negative is more costly than a false positive.

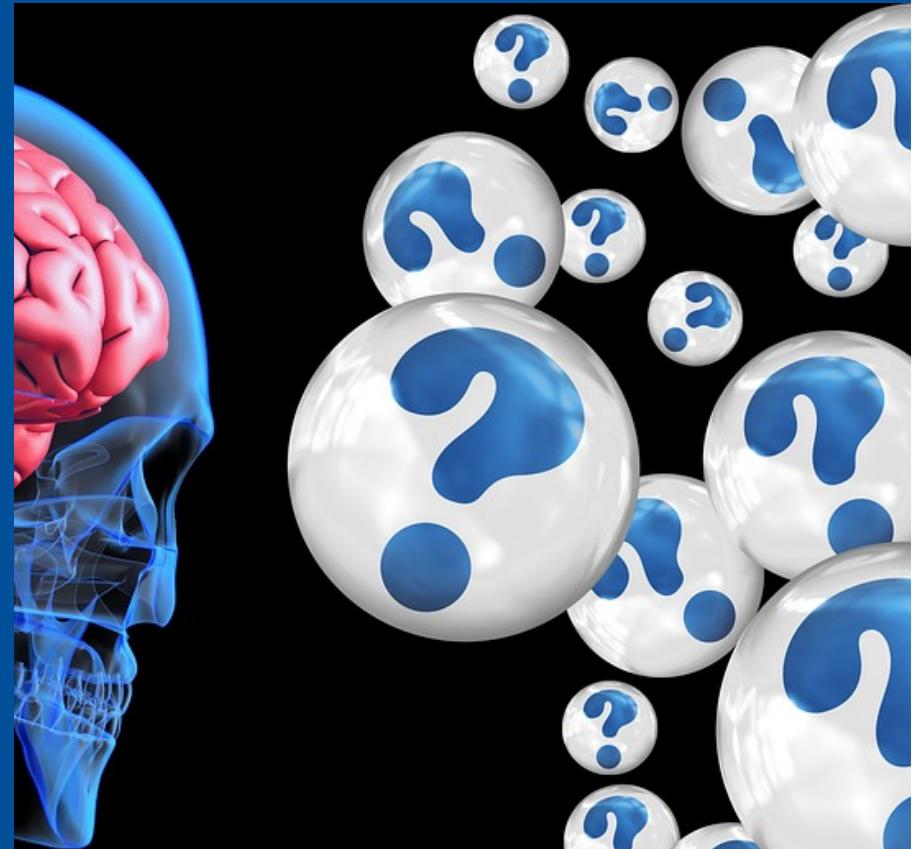
✓ BETTER THAN BAYES?

Had best recall was our Gaussian Naive Bayes model with a recall score of 94% but a false positive rate of 0.60.

Conclusions / Next Steps

Due to limitations in the dataset, we had no information about:

- Cholesterol
- Family history of stroke
- Number of strokes
- Race (i.e., the likelihood of stroke among different races)





Contact Us



TROY HENDRICKSON

troykhendrickson@gmail.com
github.com/tkhendrix22
linkedin.com/in/troy-hendrickson



JORDANA TEPPER



VIKTORIA SZONTAGH