

A decorative graphic on the left side of the slide consists of two overlapping parallelograms. The front one is blue and the back one is a light green. They are positioned diagonally, with the blue one partially covering the green one.

# Stroke Prediction

Zach Hanson



# Project Description

The goal of this project is to be able to predict whether or not a patient is at risk for a stroke based on various factors.

## What

Strokes can be caused by a blockage in the arteries leading to the brain or a blood vessel that ruptures in the brain.

## Why

According to the CDC, over 795,000 people in the United States suffer from a stroke every year, with about 610,000 of these being the first time the person has suffered from one.



# Dataset

Variable Name	Description
Gender	Whether the person was male, female, or other
Age	Age of patient
hypertension	Whether or not the patient has hypertension: 0 for no, 1 for yes
heart_disease	Whether or not the patient has a heart disease: 0 for no, 1 for yes
ever_married	Whether or not the patient was ever married
work_type	Type of work the patient was involved in (children, government job, self-employed, never worked, or private)
Residence_type	Whether the patient lived in Rural or Urban area
avg_glucose_level	Average glucose level of the patient
bmi	Body Mass Index of patient
smoking_status	If the patient currently smokes, formerly smoked, never smoked, or unknown

- Data set was acquired from [Stroke Prediction Dataset](#) on Kaggle by Fedesoriano
- This data uses 10 features to try to predict whether or not a patient is at risk for a stroke
- 5,110 Respondents



# Stakeholders

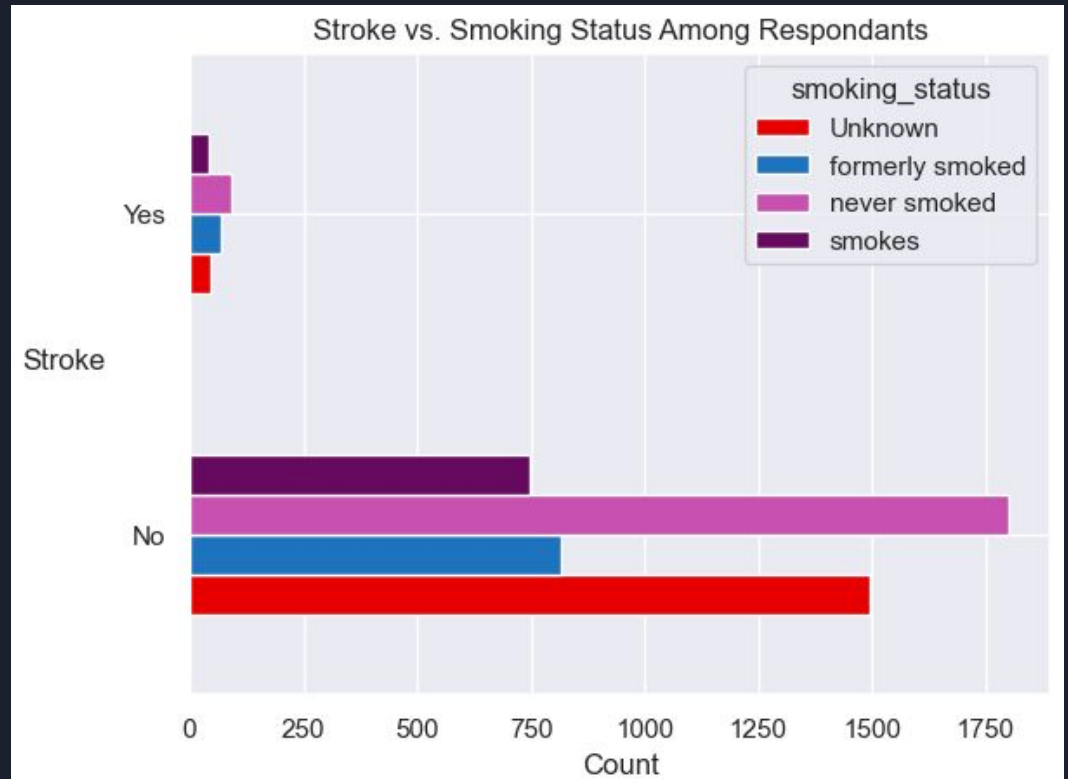
This model will be used mainly by primary care physicians in their practices to help predict whether or not their patients are at risk for a stroke.

Preventing strokes can help to keep costs down for patients, as the cost for stroke can range from \$18,000 to \$40,000 or higher in some cases depending on the type of stroke.

Some features used to predict stroke can be changed with lifestyle choices, which could lead to overall healthier patients and less risk of other health complications.

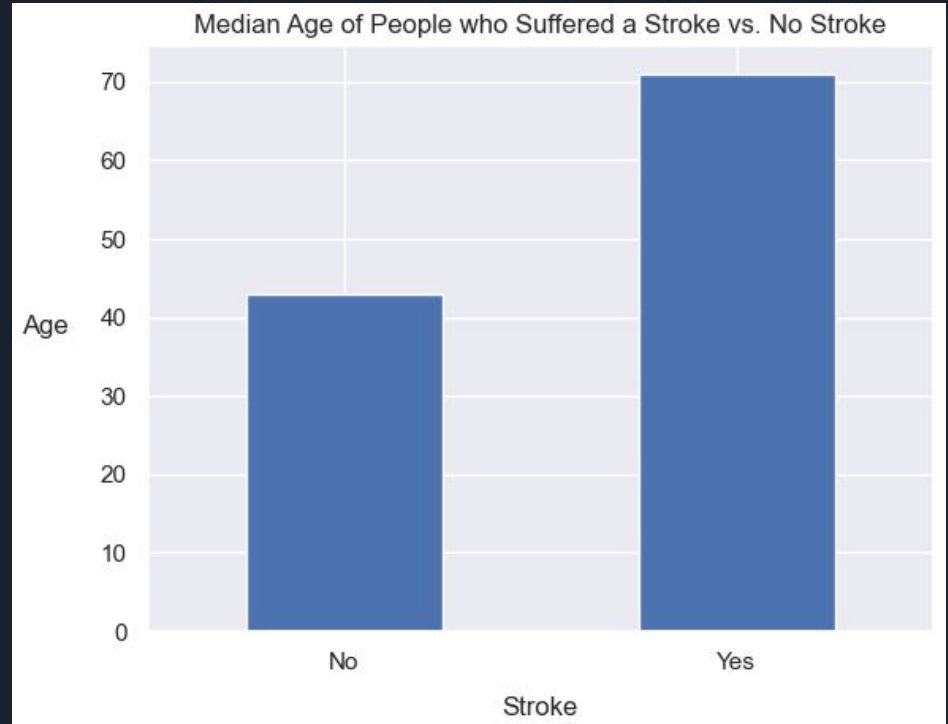
# Key Findings

- Risk of stroke slightly increases if patient currently or formerly smoked



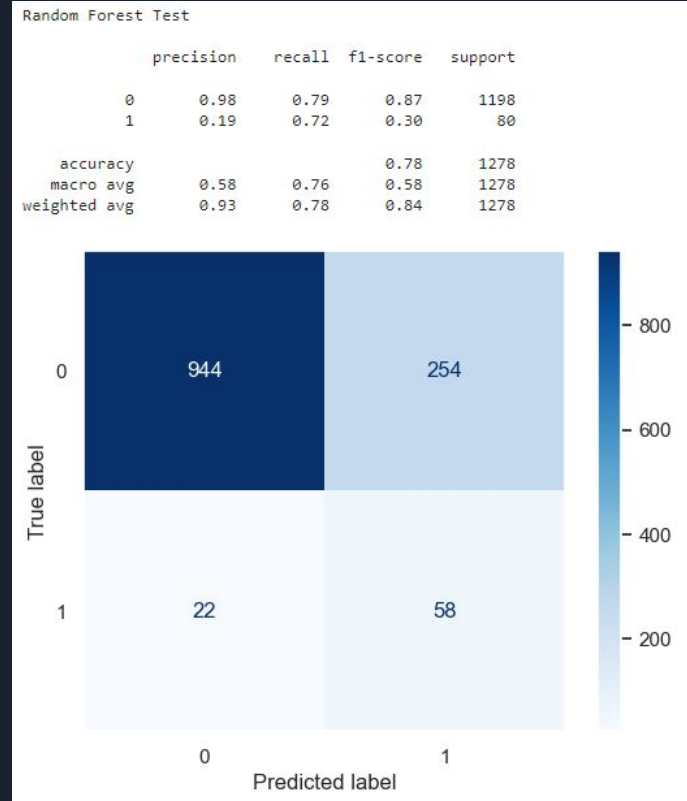
# Key Findings

- Median age higher in patients who recorded a stroke



# Model Results

- 0 means patient did not have a stroke, 1 means they had a stroke
- Want minimal number of false negatives (Bottom Left Quadrant)
- False negatives could cause patients who are at risk for a stroke to be overlooked and not receive needed preventative care





# Recommendations

This tuned Random Forest model can help with seeing whether or not a patient is more at risk for a stroke to receive the appropriate care to try and prevent one from happening.

False negatives are still a small problem and some human analysis will still be required if trying to predict if a patient is at risk.

Older patients, especially those who have potentially negative lifestyle choices such as smoking or high BMI's should talk to their physician about preventative care.