

# Stroke Prediction

Thomas Bingamon

# **CONTEXT:**

- **Stroke is the 2nd leading cause of death globally.**
- **It is responsible for approximately 11% of total deaths, according to WHO(World Health Org).**

# Our Target

- Predict whether a patient is likely to get stroke.
- Focus on gender, age, various diseases, and smoking status.

# Challenges

## Challenge 1

### Unknown Smokers

- A number of patients in our study had checked off **“Unknown”** in their smoking status.
- Are they socially smoking or are they smoking habitually?

## Challenge 2

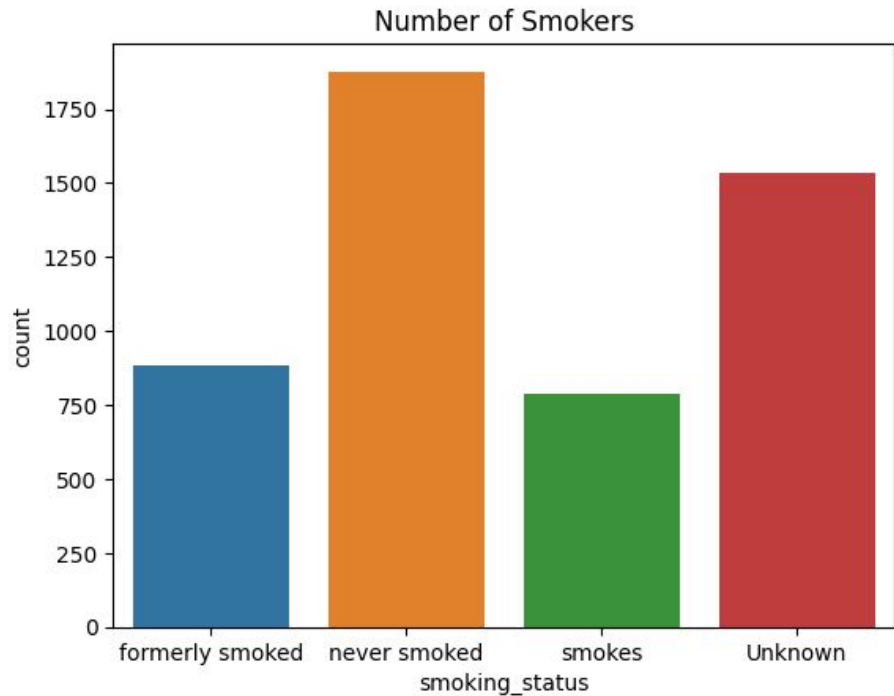
### Outliers

- Although our data was mostly revised and cleaned, there were a few outliers in our research.

## Challenge 3

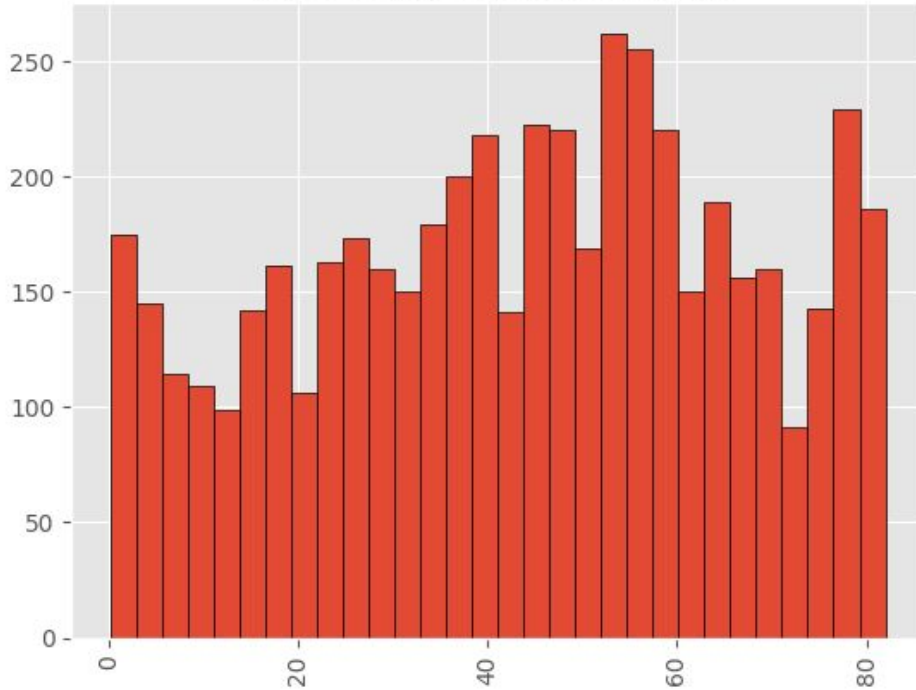
### Will our predictive models work?

- In various machine/predictive models, we were able to find a test model with high accuracy recall.



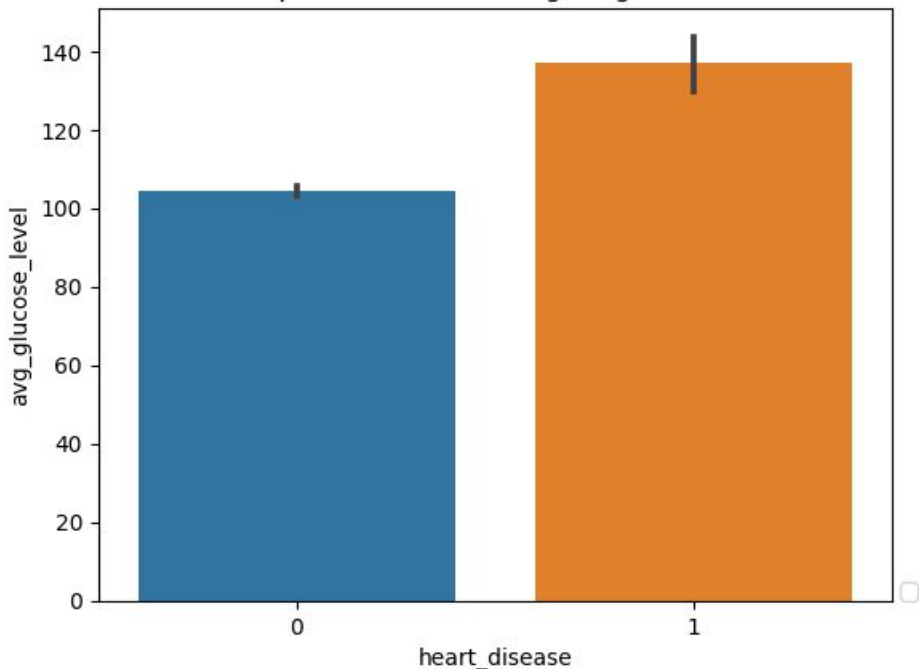
- There are a significant amount of patients who marked “unknown”.
- Patients may smoke on a social level or may be not inclined to share their true smoker status.

Age vs Likelihood of Stroke

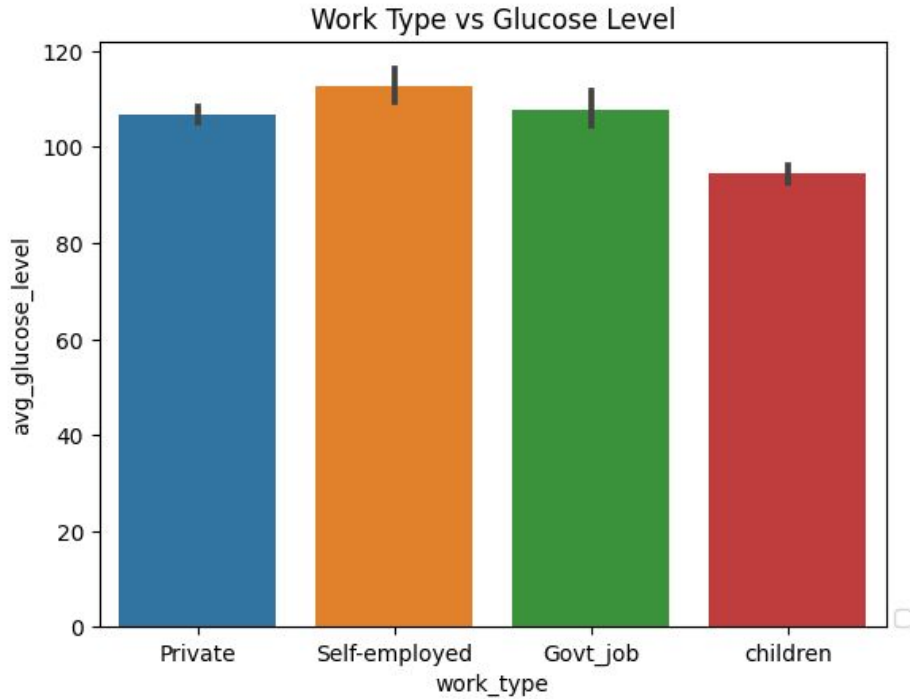


- In our findings we can see that starting from ages 50-80 there is an uptick in likelihood of getting a stroke.
- This could be for various reasons such as: higher cholesterol, glucose levels, unhealthy eating habits, less exercise, etc.

Which patients have the higher glucose level?



- Our patients labeled '0' and '1', we can see that the '1' column has an exponential amount of higher glucose average than our patients in the '0' column.
- We can delve deeper into this in the following slide.



- In a more detailed graph, we can see how even work type may have effect on one's glucose level.
- Children have the lowest glucose level.
- Self-employed have the highest.



# Implementation

- A false negative can have serious consequences, especially when positive outcomes represent critical situations(i.e. Having a stroke).
- Failure to detect positive instances result in missed opportunities, delays, or threats to safety or well-being.
- Relying on the model's predictions reduce trust and confidence if false negatives are frequent.

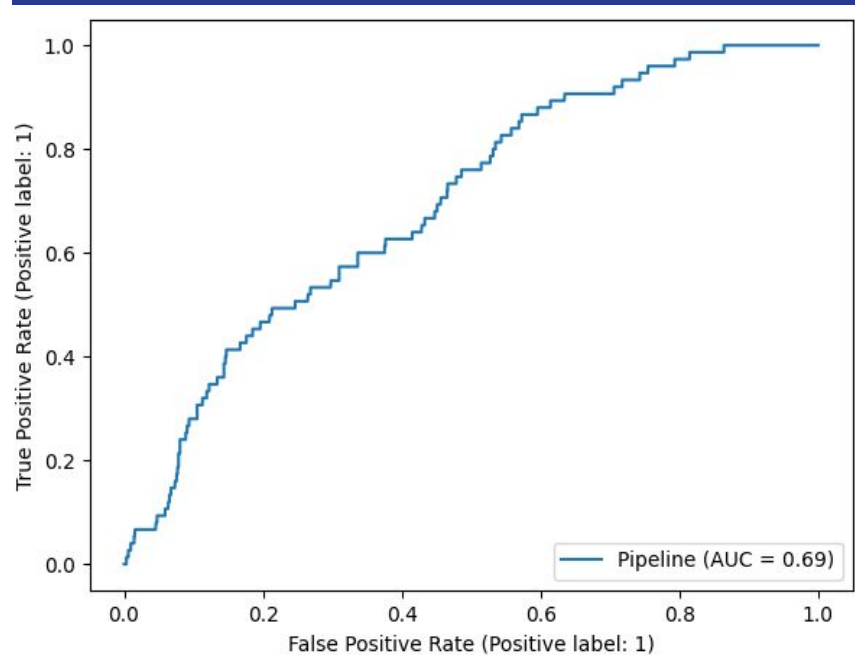
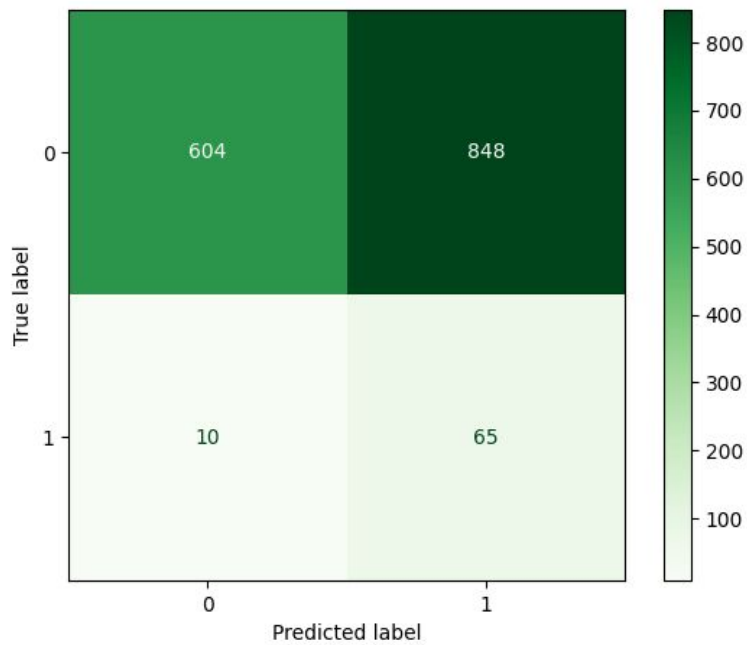
# Impact

---

- Minimizing the false negatives in our dataset is crucial.
- We should prioritize recall or sensitivity to ensure the detection of positive cases.

# Impact

---



- Through various predictive models and machine learning tools, we were able to find a suitable predictive model.
- We found a very high recall percentage of 87%, but most importantly a very low false negative.

# Impact

---

# In Conclusion:

- With a low false negative rate, positive instances are being reliably detected, reducing the need for manual review or reassessment of cases that could be false negatives.
- By decreasing missed opportunities, we can make more informed decisions based on a much deeper view of the data, leading to better outcomes and providing resources where needed.