Stroke Predictor

Analysis and Classification of Strokes

Motivation and Stakeholders

- Stroke: medical emergency caused by a lack to oxygen to the brain, often due to a blocked or ruptured artery
- Strokes are a leading cause of disability and death worldwide
- Damage/death may already be done before a treatment can be administered

Thus,

- Large benefit if medical advisors were able to predict which patients were the most at-risk for having a stroke
- Allow them to make recommendations for preventative medicines and lifestyle choices

Data Overview

Each row in this dataset corresponds to information regarding an individual patient. This included some medical and lifestyle information.

Only around 1 in 20 patients experienced a stroke.

Attribute Information

1) id: unique identifier

2) gender: "Male", "Female" or "Other"

3) age: age of the patient

4) hypertension: 0 if the patient doesn't have hypertension, 1 if the patient has hypertension

5) heart_disease: 0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease

6) ever_married: "No" or "Yes"

7) work_type: "children", "Govt_jov", "Never_worked", "Private" or "Self-employed"

8) Residence_type: "Rural" or "Urban"

9) avg_glucose_level: average glucose level in blood

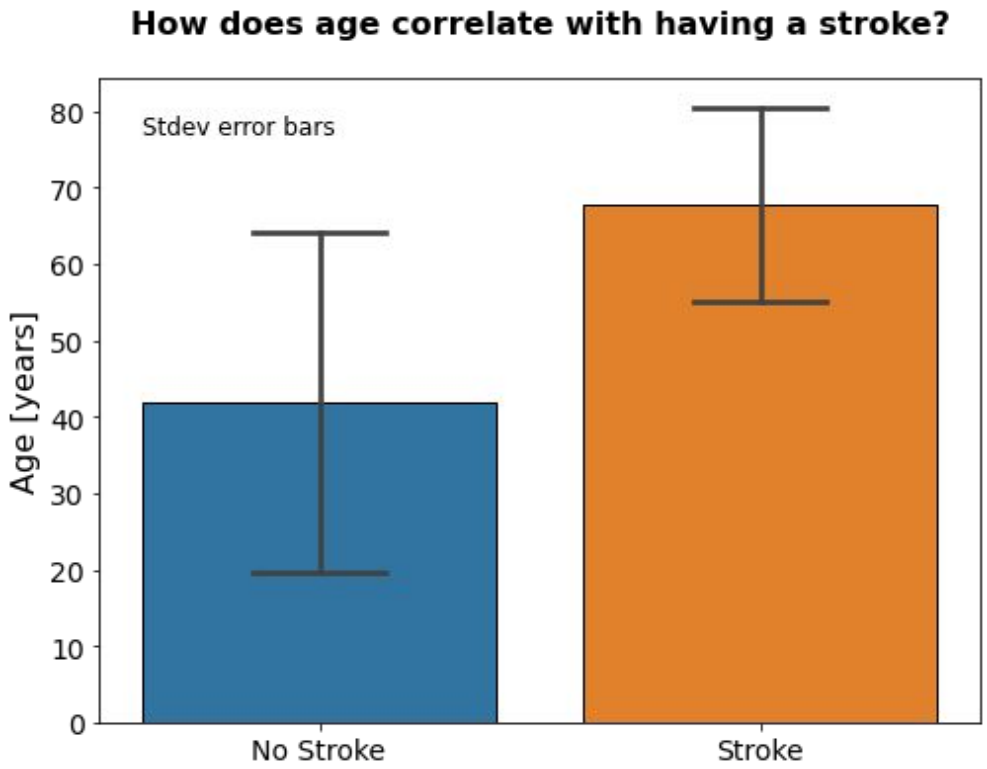
10) bmi: body mass index

11) smoking_status: "formerly smoked", "never smoked", "smokes" or "Unknown"*

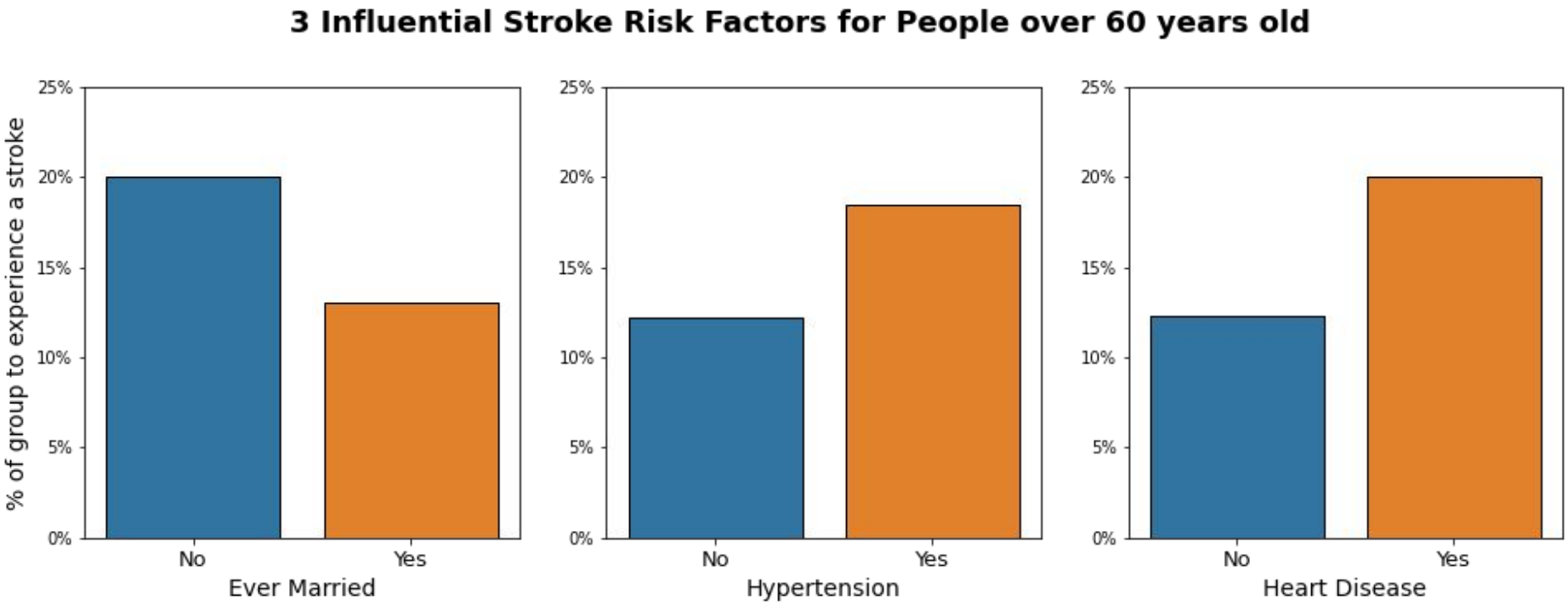
12) stroke: 1 if the patient had a stroke or 0 if not

*Note: "Unknown" in smoking_status means that the information is unavailable for this patient

Key Findings



Key Findings



Model Performance

Model developer decision:

- False negative are likely more costly than false positives

Caveat:

- Model was not able to train on many positive cases due to class imbalance

Outcome:

- Model greatly over predicts that the patient will have a stroke (see precision)
- There are still some false negative predictions (see recall)

Metric	Value
Recall	0.81
Precision	0.13
Accuracy	0.74

0 is worst, 1 is best

Recommendations

- Model predicts too many false negatives and false positives for it to be an extremely effective and high confidence tool
- Depending on the risks of the preventative medications and treatments, this model should not be used to prescribe
- That said, it could be used by a physician as an internal "at-risk" marking tool, allowing them to keep a closer eye on patients that may have a higher risk strokes
- This may led to more extensive screenings to the extremely at-risk group which could potentially save lives.