Deloitte.

Stroke Awareness Federal Campaign

Overview

Table of Contents

Our GOAL is to use MACHINE LEARNING to develop a model that can help vulnerable patients ASSESS their own risk of a stroke in lieu of traditional healthcare resources

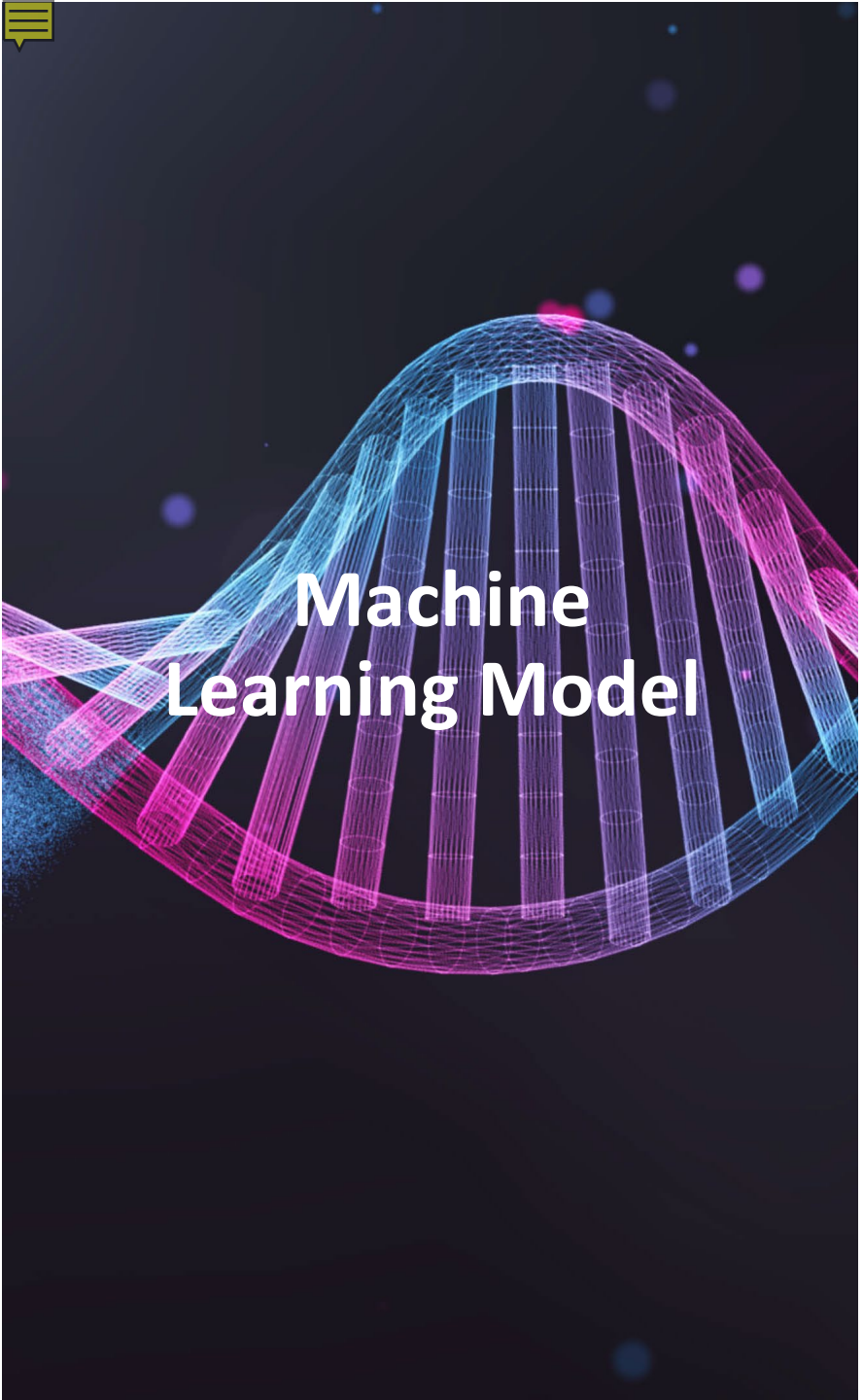
INFERENTIAL MODELING

Of the patients who were admitted to the hospital for a stroke, what factors did they have in common?

Data Modeling

PATIENT ATTRIBUTES

This slides depicts the patient attributes we will be focused on in analyzing ties between the likelihood of stroke.



Methodologies and Tools Used

- Our data has 5,109 cases, and of those 249 (**around 5%**) patients were positive for a stroke
- We used **Synthetic Minority Oversampling Techniques** (SMOTE) to address the class imbalance
- There was **no significant collinearity** within the features (<70%)
- We compared the performance of 6 different models, focusing on **recall** as our top metric

Data Modeling Process for Stroke Prediction

Our goal is to predict stroke occurrences accurately, aiding in preventative health measures.

Data Prechecks

- Ran a collinearity analysis and found no major correlation between columns
- Utilized KNN imputer to replace missing continuous values

Data Standardization

- Applied StandardScaler to normalize the data, improving the performance and stability of our models

Final Selection

- Chose the model that balanced high recall with good overall accuracy

Data Balancing Techniques

- Utilized SMOTE to balance our dataset, ensuring a fair representation of stroke cases and help prevent bias towards the majority class in predictions

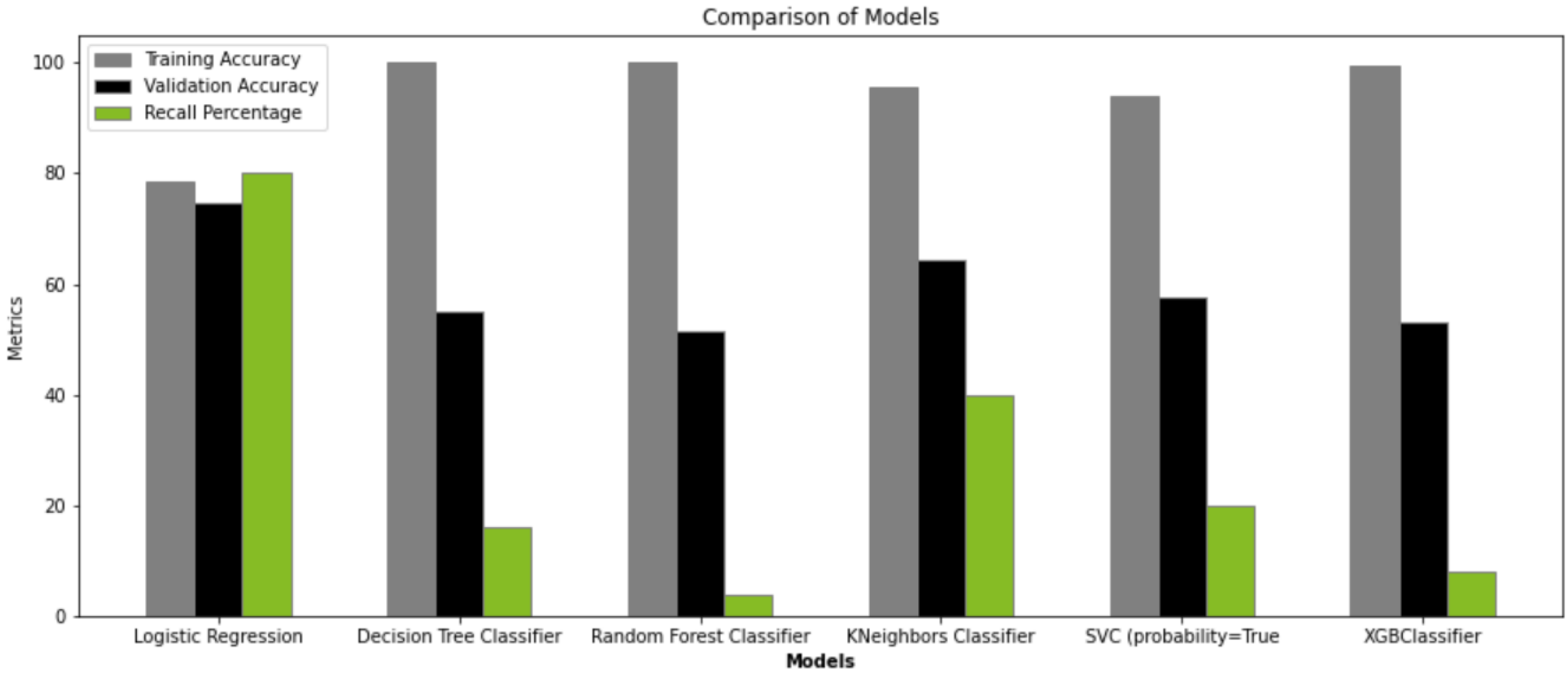
Model Training & Evaluation

- Trained and compared the performance of 6 different classification models, based on accuracy and recall

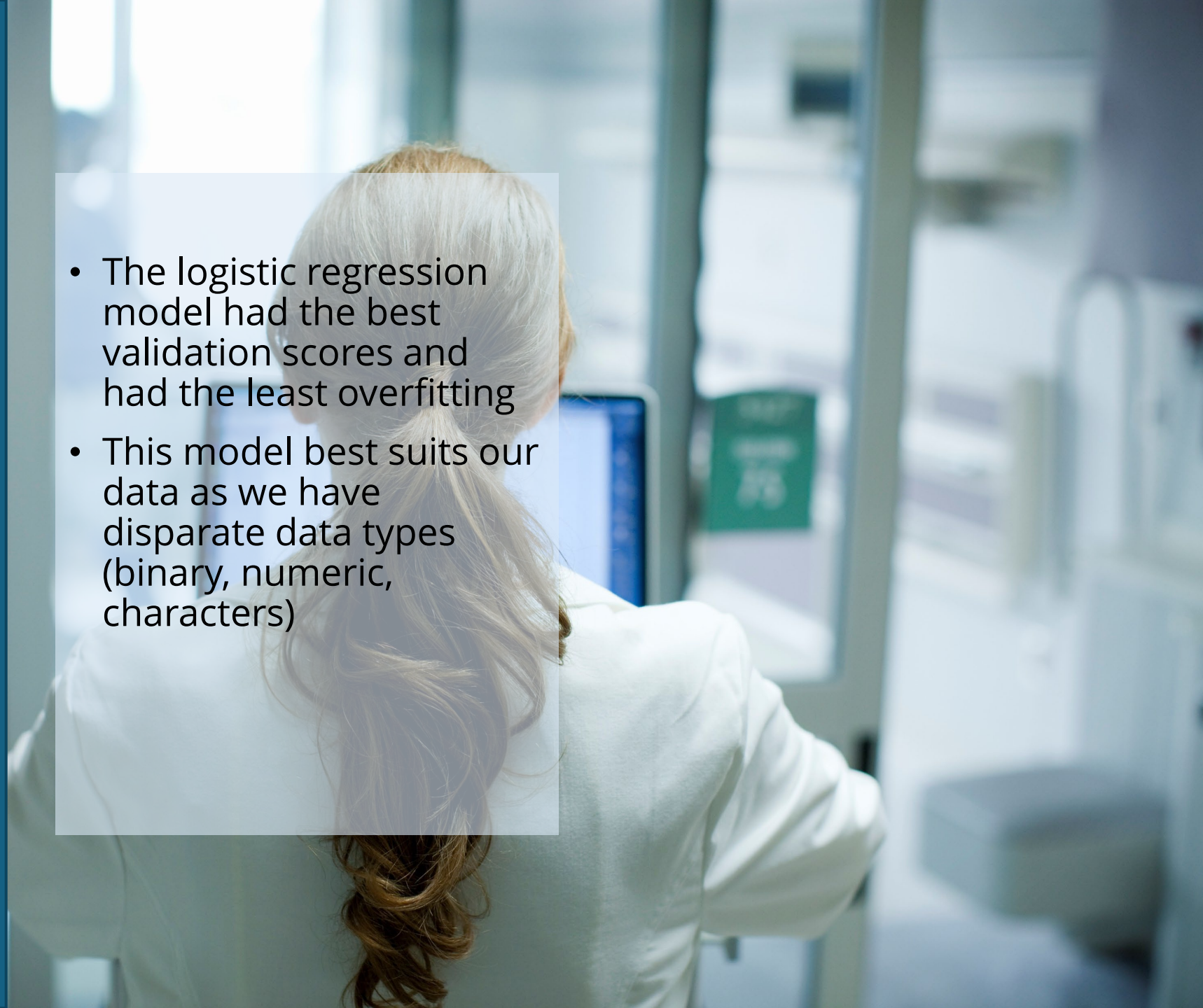
Evaluating the Model

Model Performance of Six Classification Models

Here we compared the training set accuracy, testing set accuracy, and recall. The logistic regression model had the most balanced performance and highest accuracy in predicting stroke cases (recall).



Logistic Regression Model



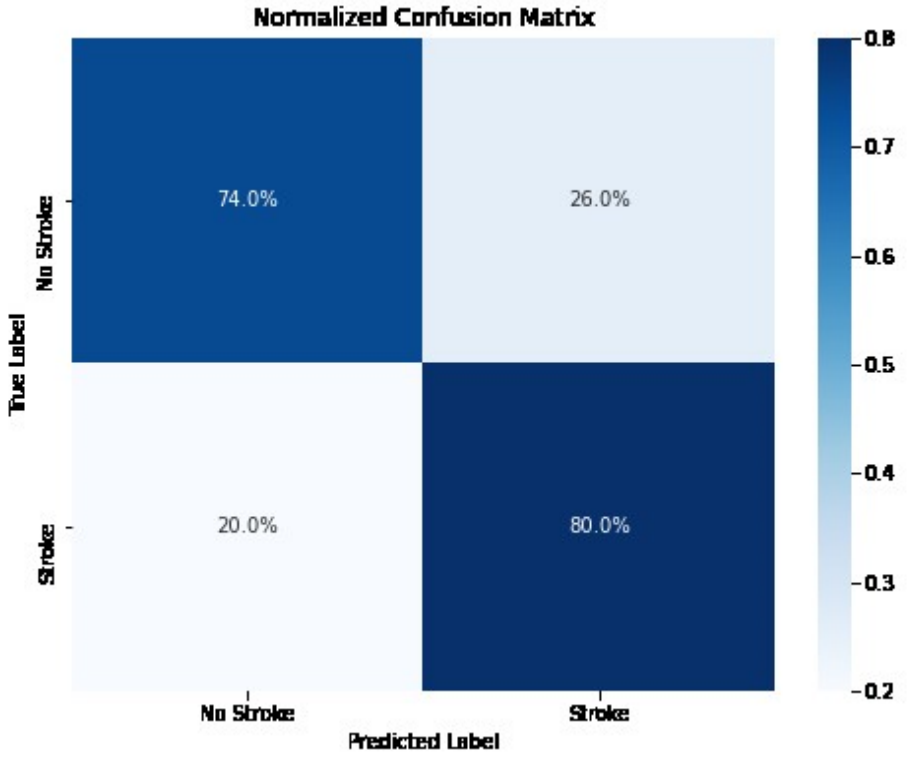
- The logistic regression model had the best validation scores and had the least overfitting
- This model best suits our data as we have disparate data types (binary, numeric, characters)

Accuracy

Our logistic regression model good had scores for **accuracy (78.5% training, 76.9% validation)** and **recall (80%)**

To improve accuracy, we included:

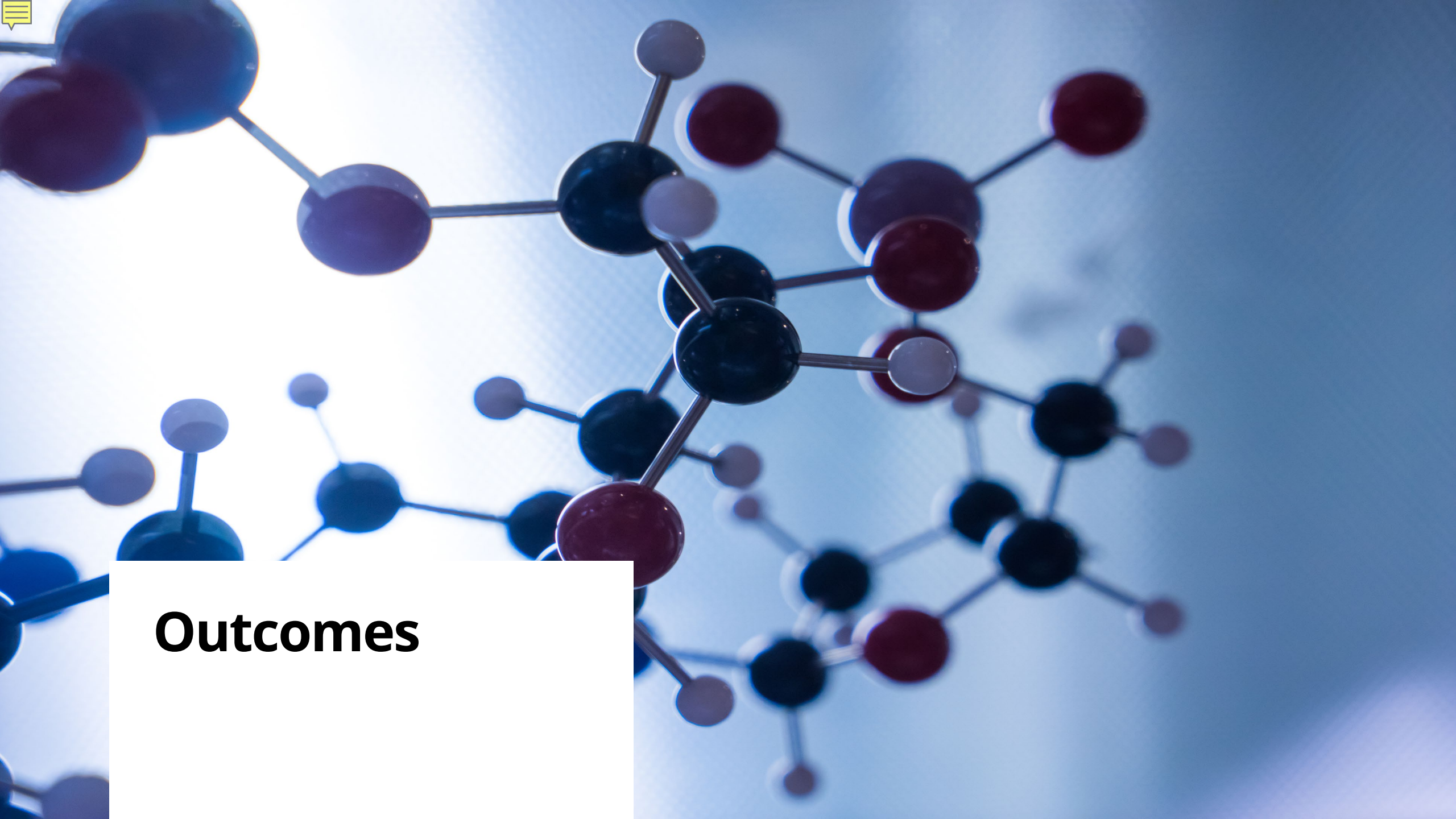
- the addition of balanced class weights to allow the model to function without being thrown off by outliers
- setting a higher recall value



Interpretation

Purpose of this model:

Using the results, we can bring awareness to people with limited access to healthcare that have certain symptoms or are identified as ‘high risk’ by their doctors

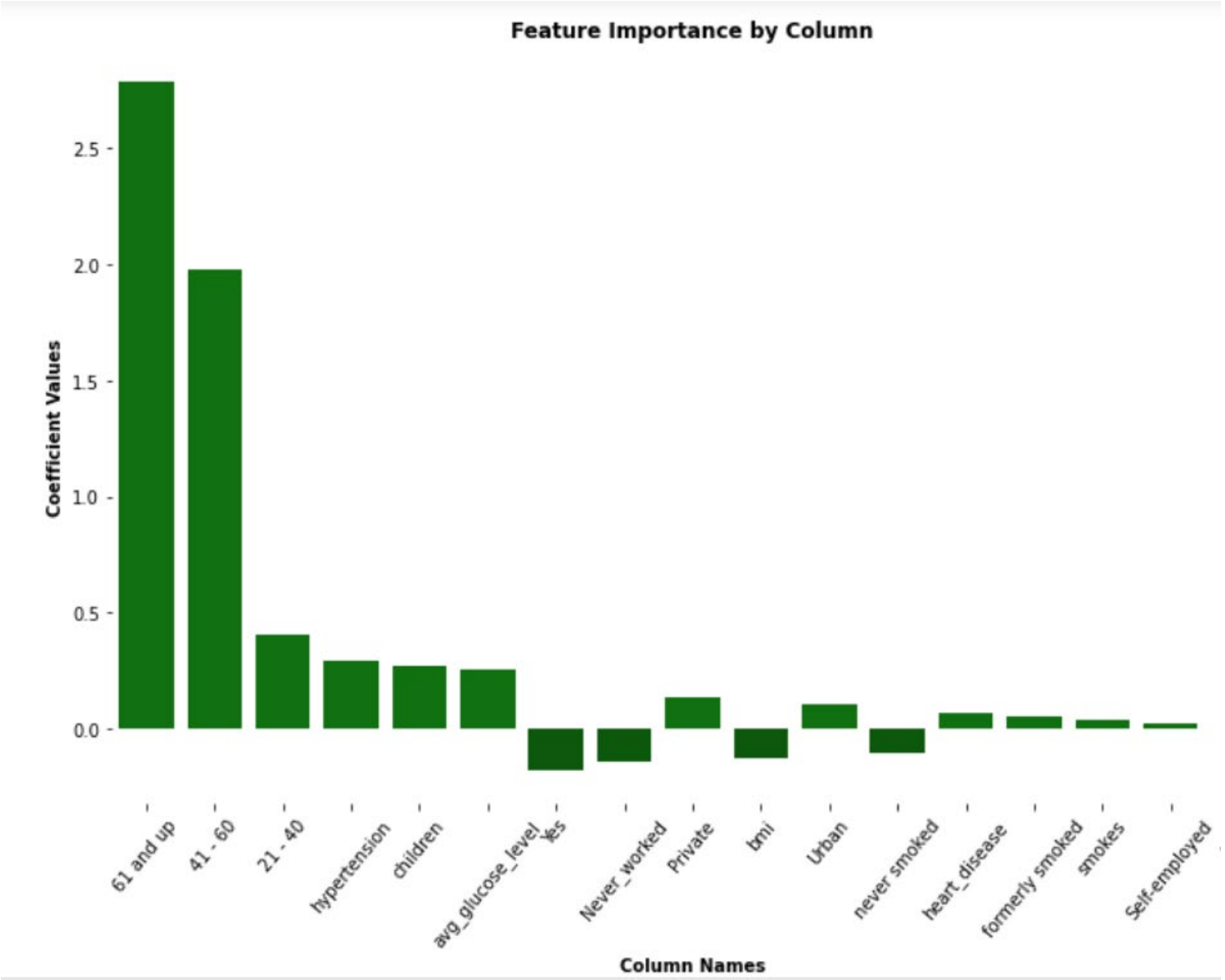


Outcomes

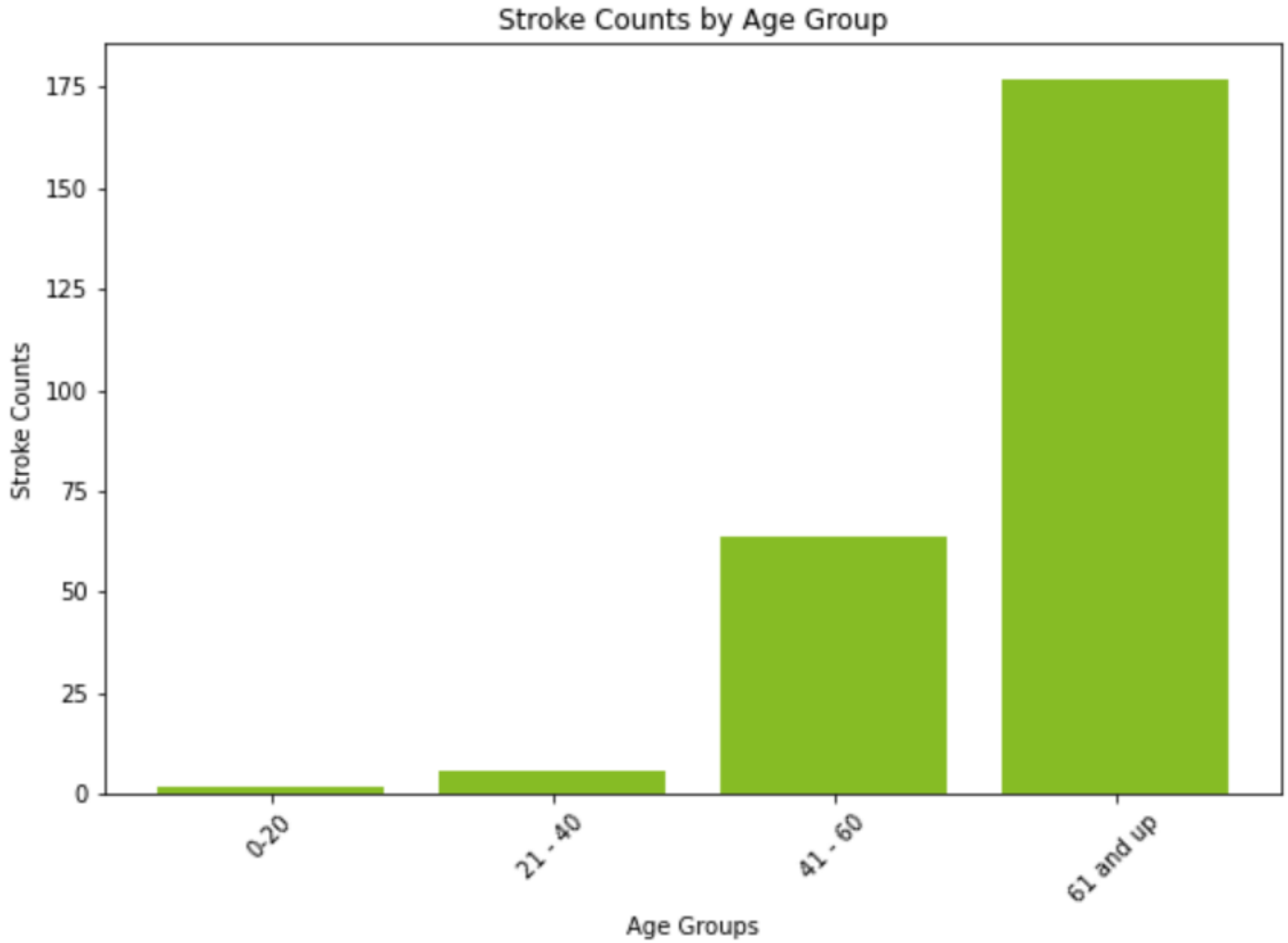
Feature Importance

Top 3 features

- Age (above the age of 41)
- Hypertension
- Children



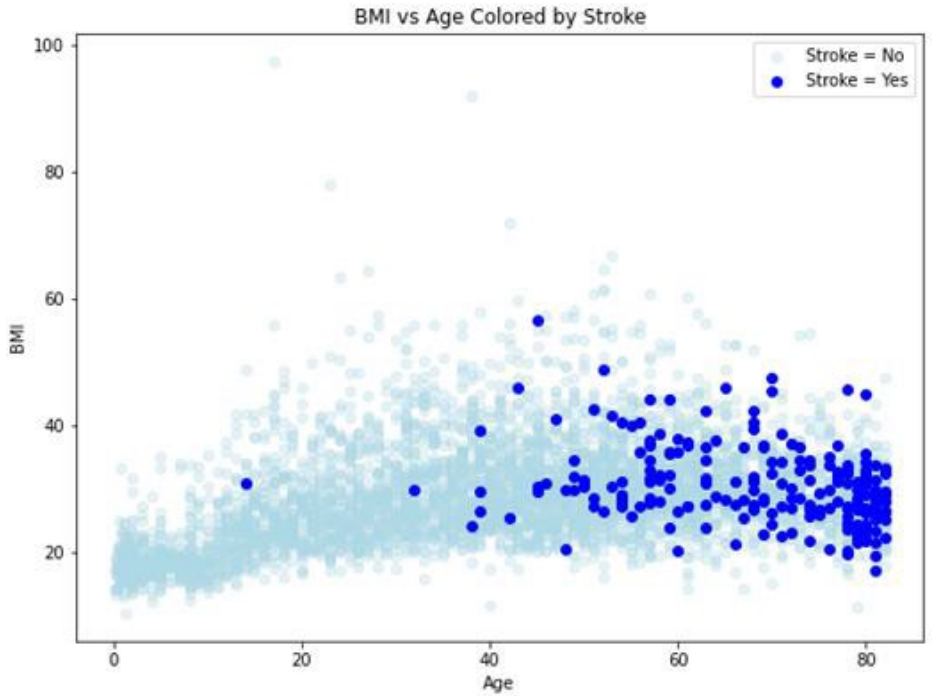
Age Distribution for Patients

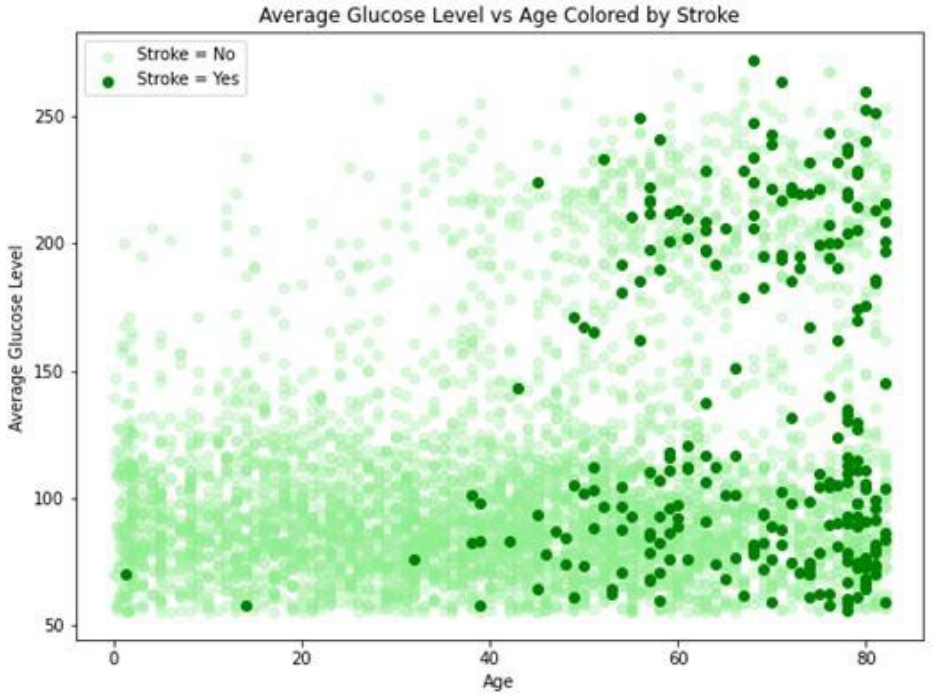


Age Factor Analysis

Here we looked at some numeric features, and we decided to do a deeper on age in relation to strokes.

- BMI and average glucose levels show a correlation with stroke incidence.
- Older age groups tend to have a higher incidence of stroke
- The visualizations highlight the importance of monitoring BMI and glucose levels as part of stroke risk assessment.





Outreach

Recommendations

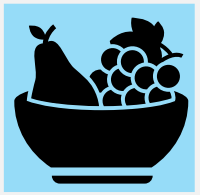
Communicate with Physicians and Primary care doctors to educate their patients above the age of 40 on stroke awareness

Build a public dashboard on stroke awareness and the high-risk factors based on our feature importances

Future Work

What factors can decrease the risk of strokes

Exercise/Activity Level



Diet and Supplements

Thank you

Thank you for listening to our proposal. Please do not hesitate to reach out to any one of us regarding any comments or queries you may have.

Connect with the Team

Alivia Conway

Email:

alconway@deloitte.com

LinkedIn:

Alivia Conway

Ellis Hartley

Email:

elhartley@deloitte.com

LinkedIn:

Ellis Hartley

Kevin Hunter

Email:

kevhunter@deloitte.com

LinkedIn:

Kevin Hunter

Donnie Lawson

Email:

dolawson@deloitte.com

LinkedIn:

Donnie Lawson

Vaia Mavilla

Email:

vmavilla@deloitte.com

LinkedIn:

Vaia Mavilla

Skylar Miller

Email:

skmiller@deloitte.com

LinkedIn:

Skylar Miller