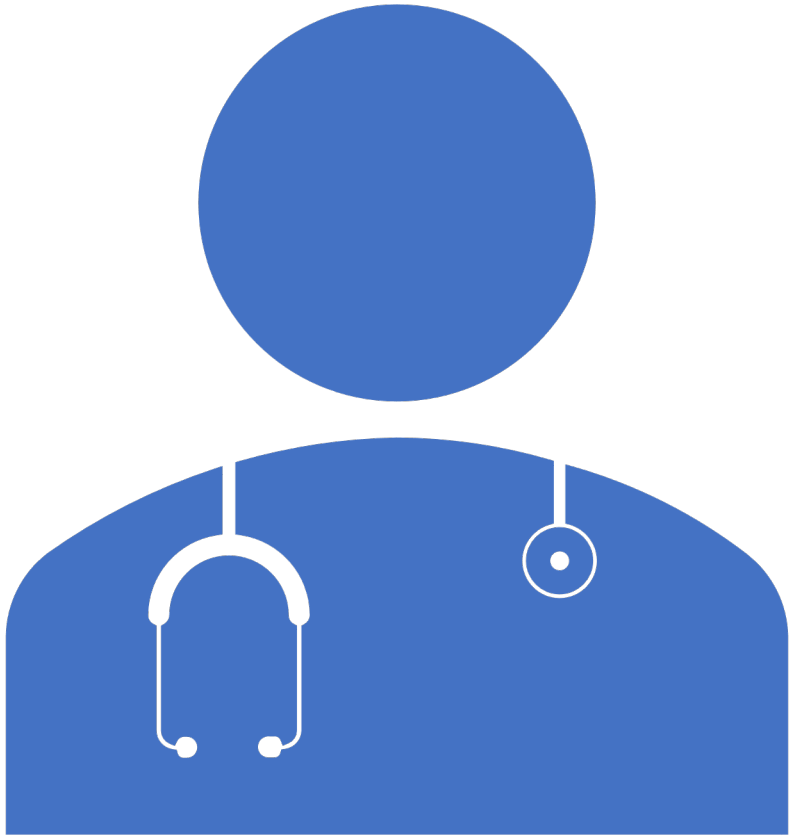
Stroke Prediction ML Model

David Patton

Business Case

- A healthcare company wants to be able to predict a stroke episode based on a variety of features.



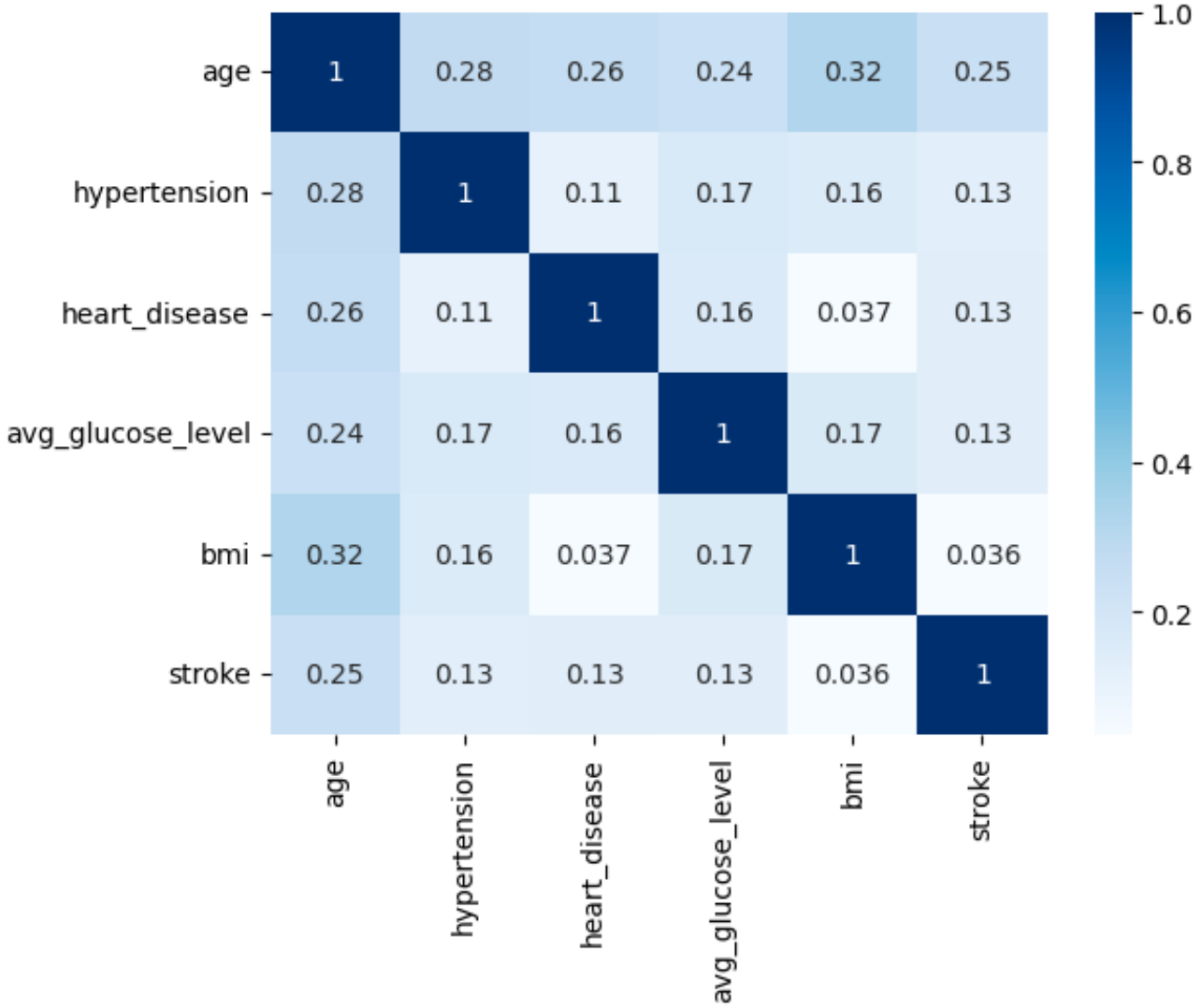
Data Info.

The dataset uses personal information and medical history about the patients.

Features: gender, age, hypertension, heart disease, marital status, work type, residence type, average glucose level, bmi, and amoking status.

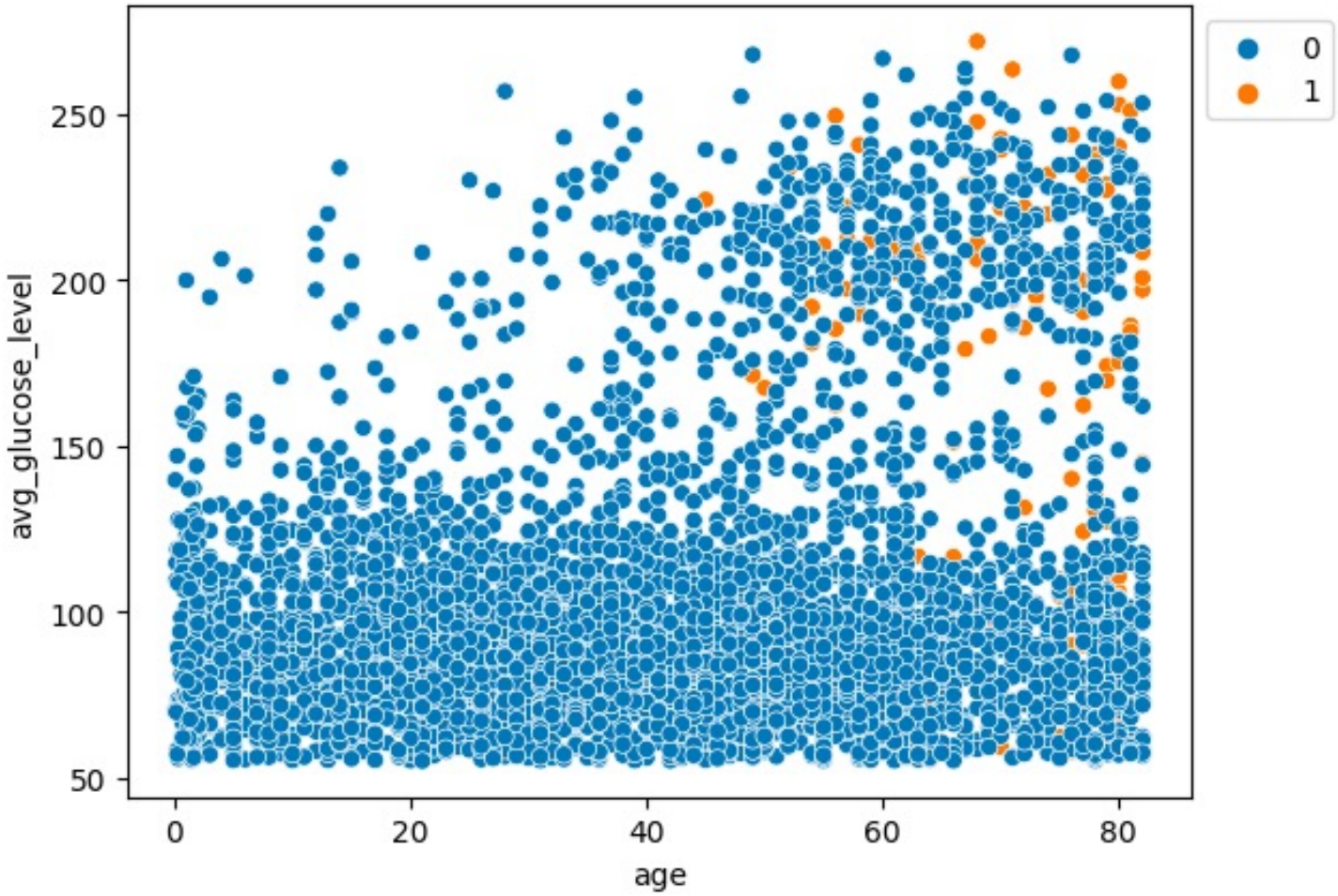
Data Visualizations - Heat Map

- The age feature has some descent correlations, especially with bmi and hypertension. The rest of the coorelations are unfavorable
-



Data Visualizations – Scatter Plot

- Looking at the scatterplot above I can conclude that patients 40 years old and up are at higher risk of stroke. However, there is no real coorelation between glucose level and age.
-



XG Boost Model (SMOTE) - Performance

- The XG Boost model was the best performer out of the models tested. I used the SMOTE technique to help correct some of the bias in the data. The f1-score for the test is 11% for stroke prediction and 96% for a non stroke preiction. Out of the models tested this is the only model I would recommend.
-

```
======Train Set Metrics======
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      3628
           1       0.98      0.93      0.95       187

    accuracy                           1.00      3815
   macro avg       0.99      0.96      0.98      3815
weighted avg       1.00      1.00      1.00      3815

======Test Set Metrics======
              precision    recall  f1-score   support

           0       0.95      0.97      0.96      1210
           1       0.14      0.10      0.11        62

    accuracy                           0.93      1272
   macro avg       0.55      0.53      0.54      1272
weighted avg       0.91      0.93      0.92      1272
```