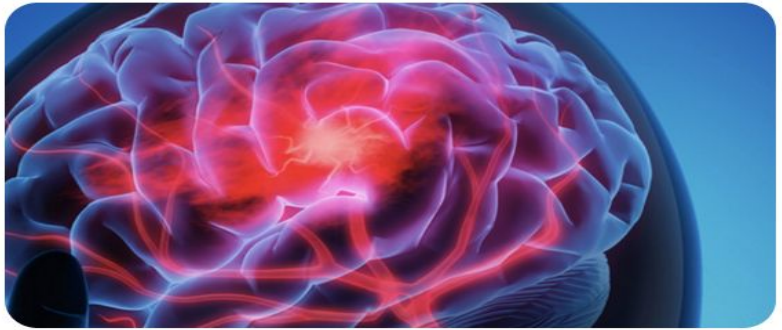


Stroke Prediction App

Using Streamlit for Predictive Healthcare

Team: 05

Motivation

- According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths.
- This dataset is used to predict whether a patient is likely to get stroke based on the input parameters like gender, age, various diseases, and smoking status.
- Our target users are mainly **Doctors**

Objectives

- **Predict Stroke Likelihood:** Use key health data to estimate stroke risk.
- **User-Friendly Interface:** Provide an intuitive platform for users.
- **Personalized Assessments:** Deliver personalized risk evaluations to empower proactive health management.

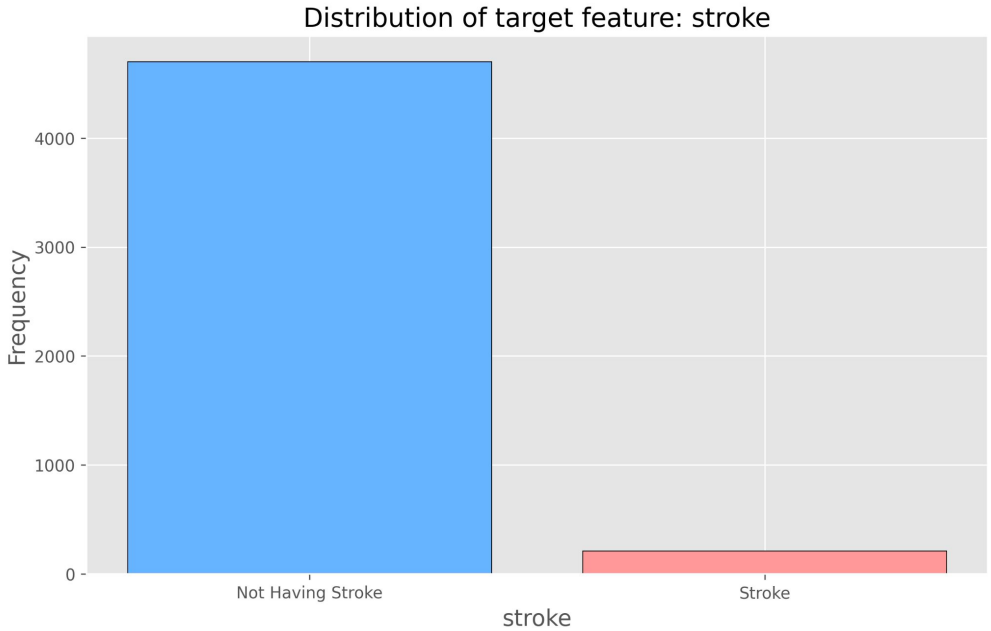
About Dataset

- Dataset is collected from Kaggle
- Dataset contains
  - 5110 patient data
  - 12 features

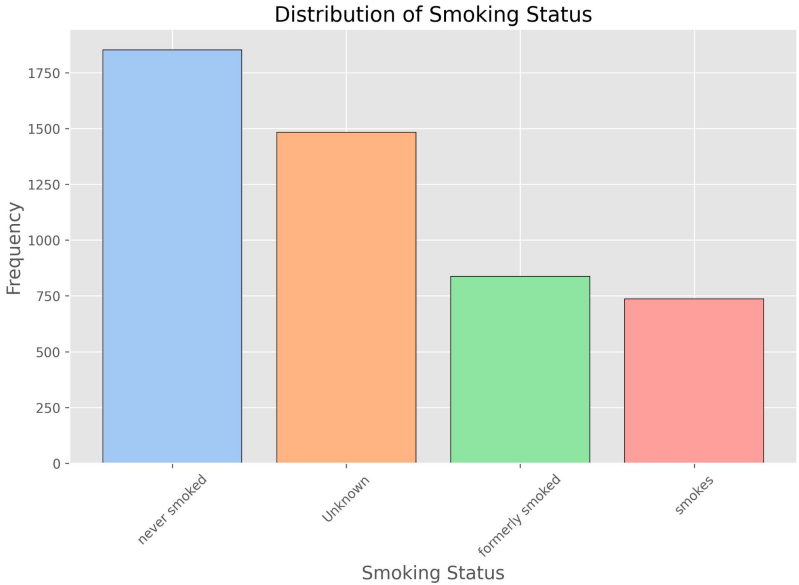
Total 12 Feature
Target Column : Stroke

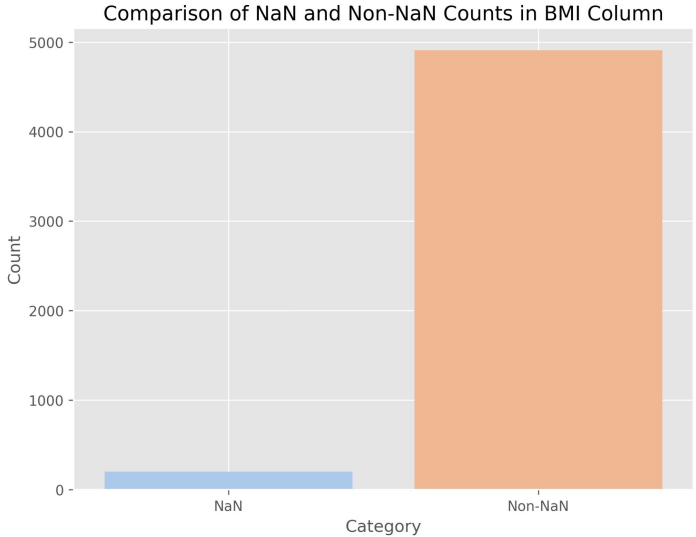
Imbalanced Target Feature

- Fewer stroke cases than non-stroke cases
- SMOTE Re-Sampling



Data preprocessing

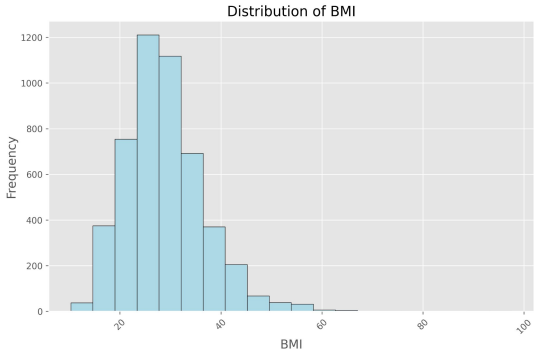


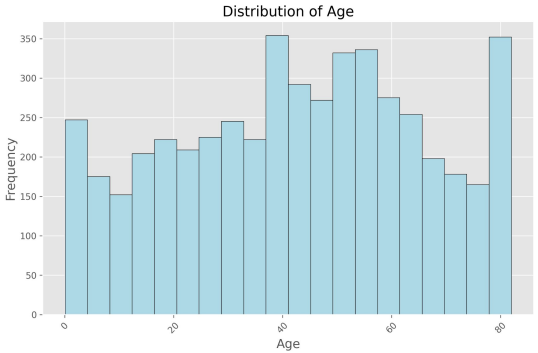


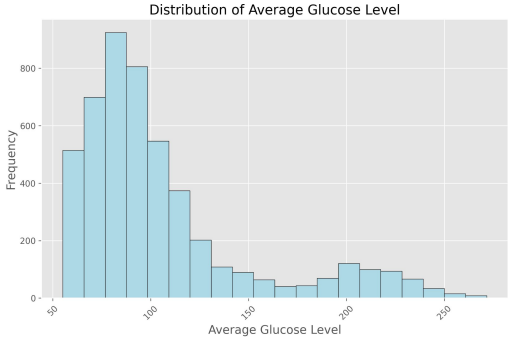
- Formerly smoked -> smokers
- Removing ‘unknown’

- Removed NaN

Data preprocessing (Cntd.)



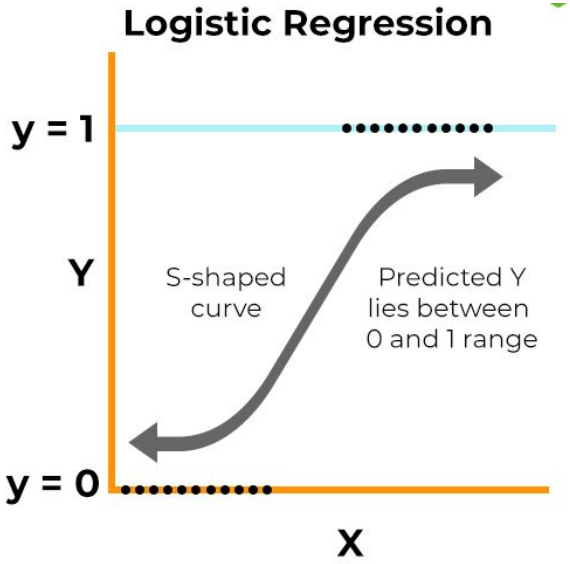




- Standardized the continuous features

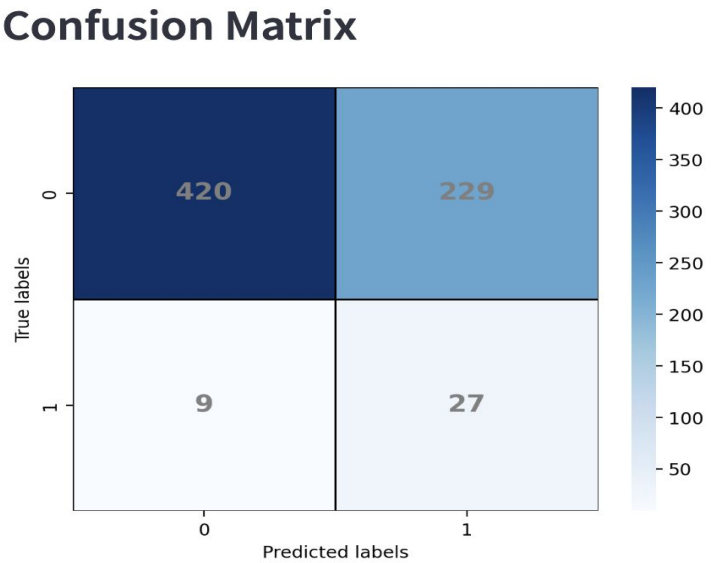
Model

- Number of data points used for training: 3425
- Logistic regression model



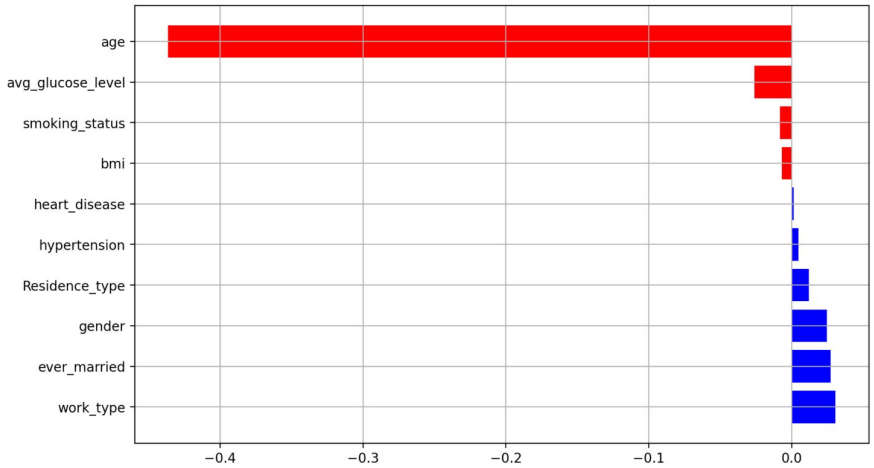
Confusion Matrix & ROC Curve

- Minimized the False Negative.
- Resulted in moderate False positives
- Threshold = 0.35



Feature importance analysis

- Shapely values
- In the plot:
  - Magnitude: The size of the bar indicates the strength of the feature's influence on the prediction. Larger bars have a more significant impact.
  - Direction: The color coding (and sign) of the bars indicates whether the influence is positive (blue) or negative (red).



Questions Addressed

1. What are the key risk factors for stroke?
2. When and when not to rely on AI’s prediction?
3. How accurately can these risk factors predict the likelihood of a stroke?
4. How can we make this prediction accessible and understandable to users?

Design Choices

User Interface (UI)

- **Streamlit:** Selected for ease of use, rapid development, and interactive widgets.
- **Layout:** Designed to structure input fields clearly and present output results effectively.

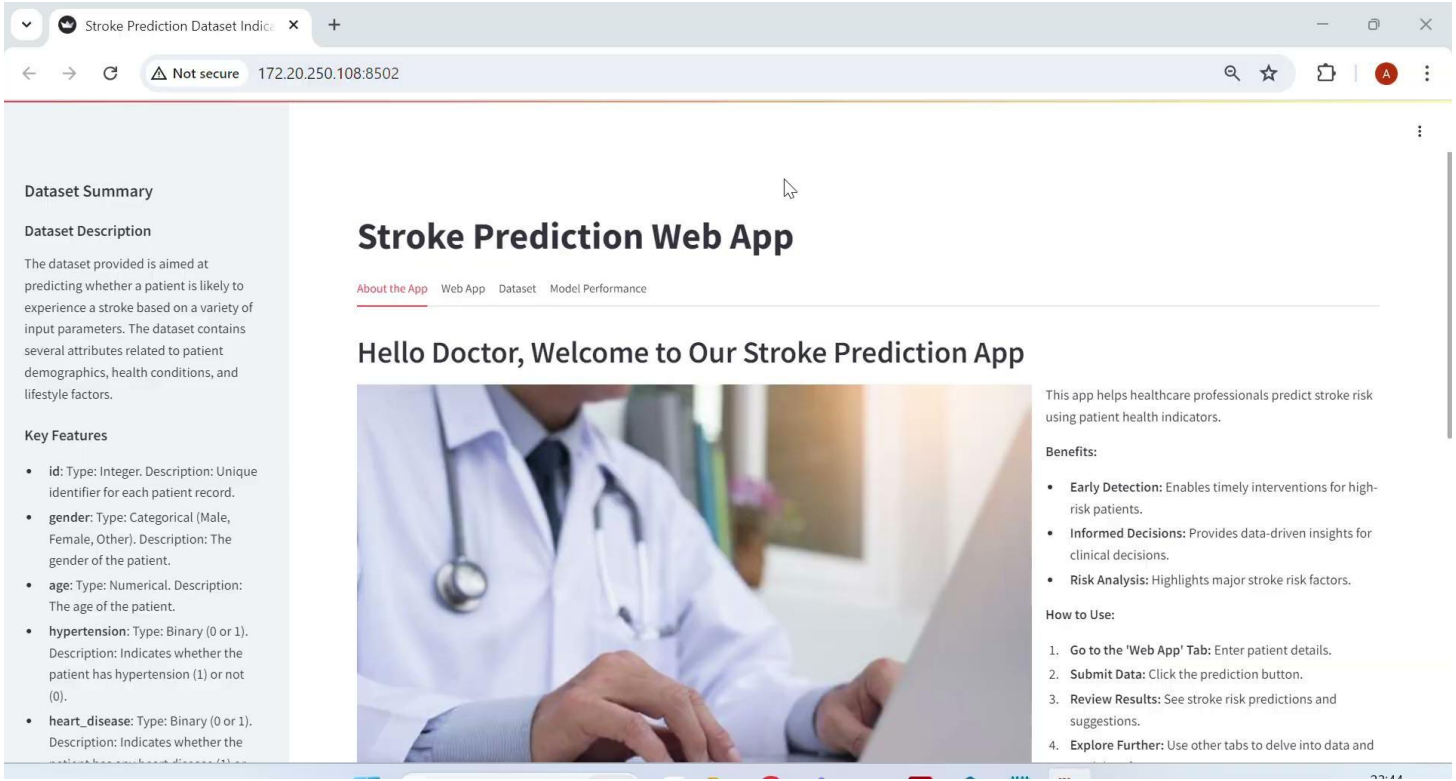
Model Selection

- **Logistic Regression**: Chosen for its simplicity, interpretability, and effectiveness with binary outcomes.

Data Handling

- **Dataset:** Utilized a comprehensive health dataset for training.
- **Preprocessing Steps:** Handled missing values and applied normalization to ensure data quality.

Live Demonstration



Feedback and Iterations

- Users appreciated the ease of use and found the predictions to be accurate.
- Improved the UI for better navigation and clarity.
- Feedback was crucial in refining the final design.
- Learned the importance of user-centered design and iterative development.

Conclusion

- Making stroke prediction simple and accessible
- Encourages people to take charge of their health.
- Early risk identification can lead to timely interventions, ultimately reducing stroke-related mortality and disability.

Thanks for listening

Appendix

Value_mappings:

```
{'gender': {0: 'Female', 1: 'Male'},
 'ever_married': {0: 'No', 1: 'Yes'},
 'work_type': {0: 'Govt_job',
  1: 'Never_worked',
  2: 'Private',
  3: 'Self-employed',
  4: 'children'},
 'Residence_type': {0: 'Rural', 1: 'Urban'},
 'smoking_status': {0: 'never smoked', 1: 'smokes'}}
```

Introduction

- Our Stroke Prediction App uses machine learning to estimate an individual's risk of having a stroke.
- It considers health parameters such as age, blood pressure, and lifestyle factors.
- Stroke prediction is crucial for early intervention, reducing mortality, and long-term disability.