Heart Stroke Risk Prediction using ML and Explainable AI

Monica Janardhan Raju, Vaishnavi Mohan

INTRODUCTION - Motivation

- Stroke remains a leading cause of death around the world.
- Stroke Prediction can be beneficial for-
  - Early Intervention and Prevention, Reducing Morbidity and Mortality
  - Optimized Resource Allocation and Cost Savings
  - Personalized Treatment for quality of life

OUR PROPOSAL

Our idea is to train models to determine the likelihood of stroke in a person based on some basic and easily available attributes.

Risk Stratification

Develop a risk stratification model that *categorizes individuals into different risk groups based on their likelihood of experiencing a stroke.*

This can help prioritize interventions and provide personalized recommendations for prevention and management.

Explainable AI

Through XAI tools *explain the behavior of the model* and help understand the impact of different features on stroke prediction .

Since interpretability is crucial in healthcare, this would provide confidence in the model's prediction by explaining its derivation.

DATASET OVERVIEW

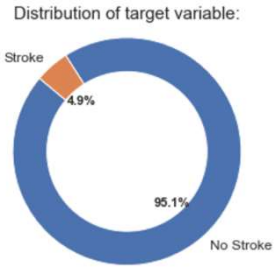
The dataset used for this project contains information about heart stroke occurrences among multiple patients. It includes various features related to **id, gender, age, hypertension, heart disease, average glucose level, BMI, smoking status, marriage status, work and residence type.**

Imbalance in data:

- The dataset is imbalanced
- ~5% patients have experienced stroke
- Class imbalance => Underperformance

Oversampling:

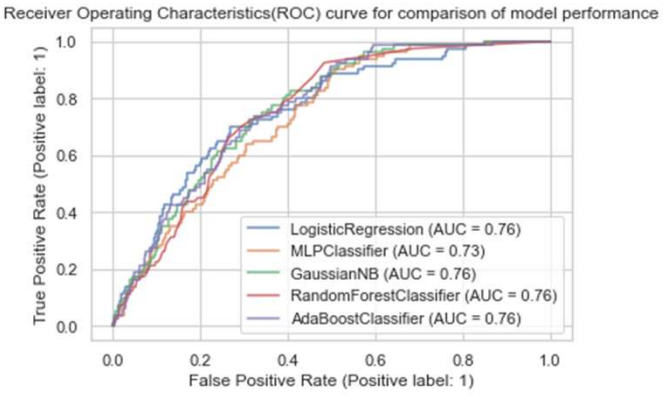
SMOTE technique is used to overcome the Class imbalance problem.



ML models: Fitting, Evaluation and Training

The following models have been implemented using Scikit Learn:

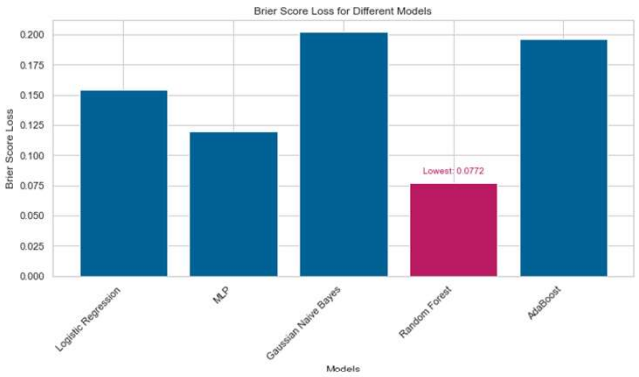
1. Logistic Regression
2. Multi layer perceptron (MLP)
3. Gaussian Naive Bayes
4. Random Forest Classifier (selected as the best-performing model)
5. AdaBoost classifier

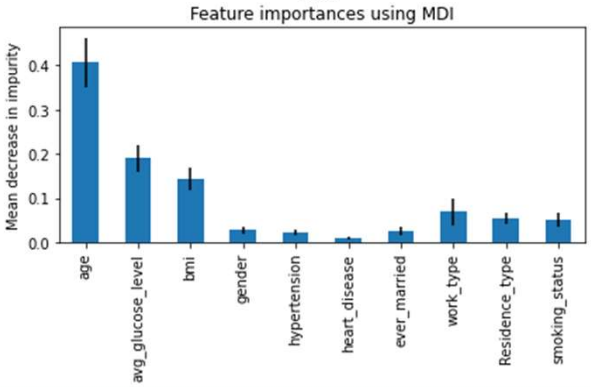


Since the AUC was similar for all selected models, Brier Scores were calculated to select Random Forest as the best performing model, as it had the lowest Brier score, significantly lesser than others. The below are its performance metrics:

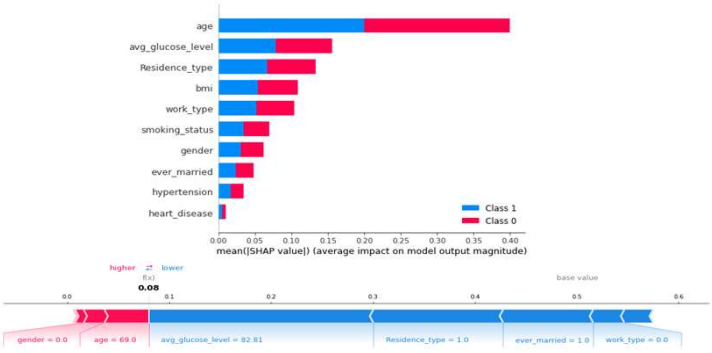
- **Accuracy – 0.9006**
- **Precision - 0.1215**
- **Recall – 0.1625**
- **F1-score – 0.1390**

MODEL METRICS & FEATURE RANKING

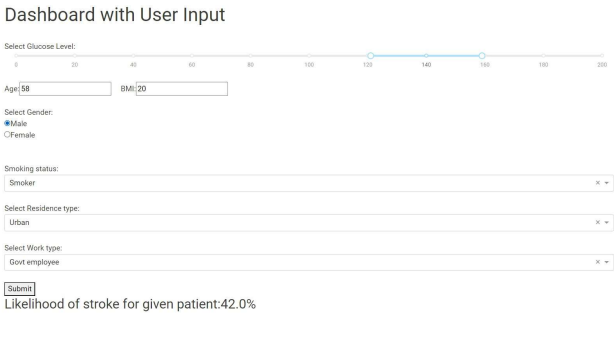




EXPLAINABLE AI



INTERACTIVE DASHBOARD



CONCLUSION

Through meticulous analysis of a comprehensive dataset encompassing vital patient attributes, the project successfully developed and evaluated a range of machine learning models. Among these, the Random Forest Classifier emerged as the most adept at discerning stroke risk patterns. It can be inferred that stroke risk increases with age and is also significantly affected by blood glucose levels and BMI.

www.PosterPresentations.com