Heart Attack Prediction Model

Will Wu

Feb 28th, 2025

Problem Framing – Heart Attack

Why

- **Personal experience**
- **805,000 heart attacks occur annually in the U.S.**

What

Early Detection System

- Preventive medications
- Awareness of relevant symptom

Drawbacks in Existing Models

- Age range 30-79
- Need blood test

How

Classification

- Predict risk of heart attack (YES/NO)

Accessible Inputs

- Basic information, lifestyle
- Accept all ages

Even 1% Reduction in Heart Attack

PEOPLE

8050

Individuals in U.S. change their lives annually

MONEY

161

Million USD saved from healthcare expenditures annually

Dataset

Description

- Sourced from Kaggle, Indicators of Heart Disease (2022 UPDATE)
- Originates from CDC's Behavioral Risk Factor Surveillance System (BFRSS)
- Surveys over 400,000 adults across all states in U.S.

Rows	• Over 440,000
Target	• One, Binary
Feature	• 39 Features • 6 Numerical • 33 Categorical

Field of Interest

- Had heart attack

Health and Lifestyle

- Sleep hours
- Alcohol drinkers

…

Health Assessment

- Last checkup time
- HIV test

…

Basic Information

- Sex
- Age
- BMI

…

Medical History

- Had diabetes
- Had stroke

…

Vaccination and Drug Use

- Tetanus shot
- Flu vaccine

…

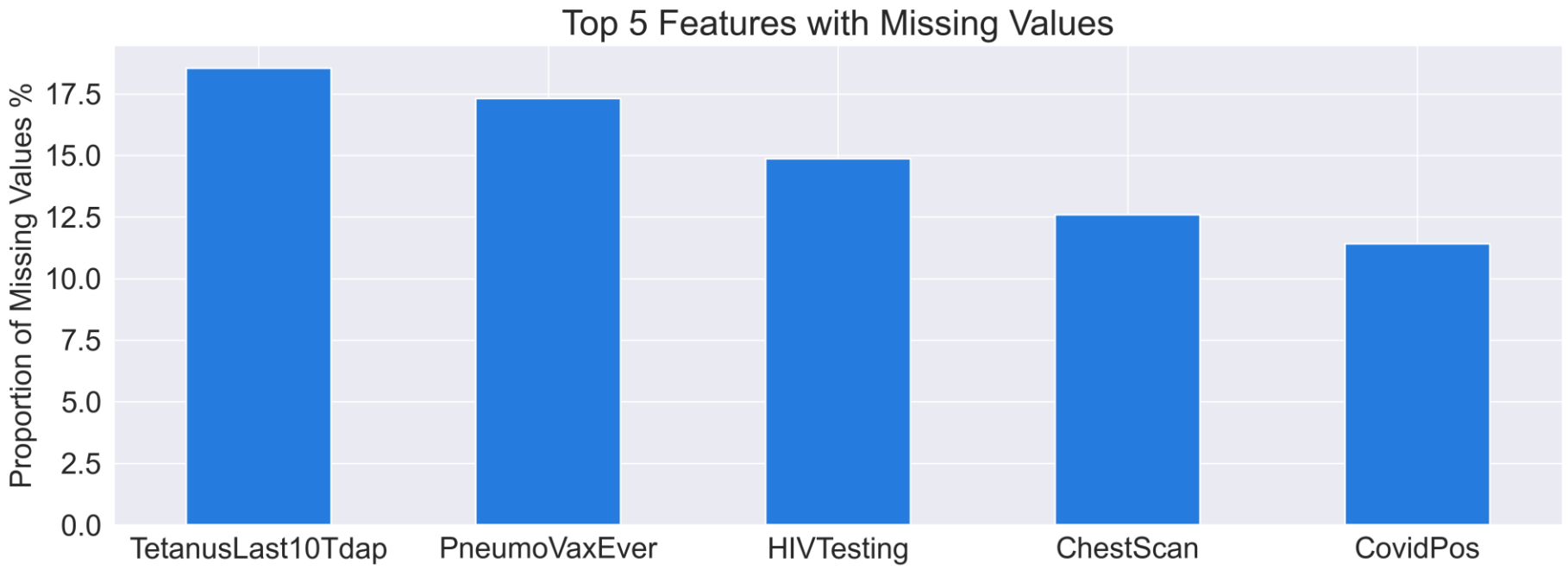
Data Quality

Imbalanced Target Variable

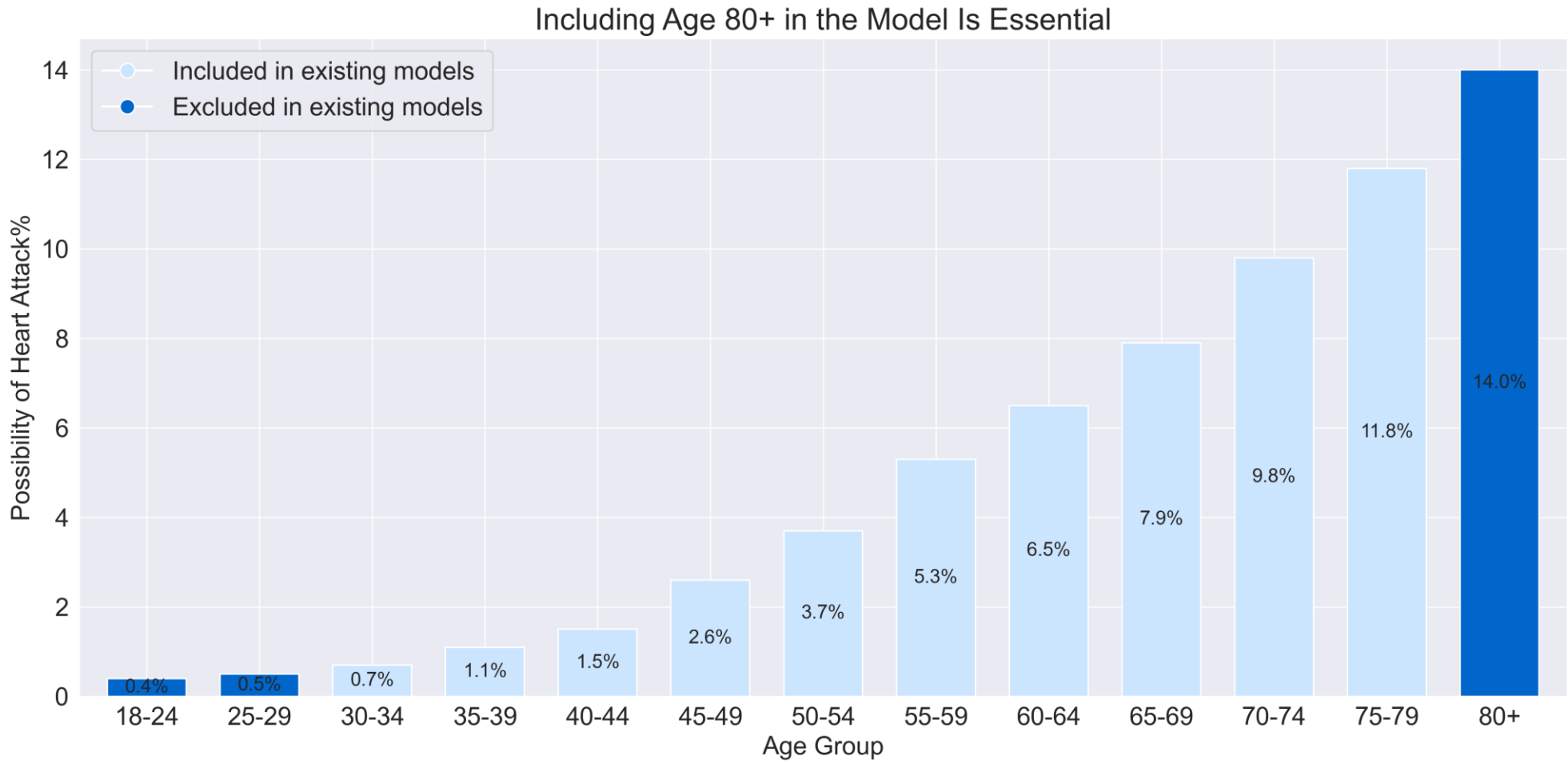
- **Only 6%** of observations had heart attack

Missing Values

- **45% of rows** contain missing values
- **95% of columns** contain missing values



Insights


Including Age 80+ in the Model Is Essential
Included in existing models
Excluded in existing models
Possibility of Heart Attack%
0
2
4
6
8
10
12
14
0.4%
0.5%
0.7%
1.1%
1.5%
2.6%
3.7%
5.3%
6.5%
7.9%
9.8%
11.8%
14.0%
18-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80+
Age Group

Next steps

1. In-depth EDA on features
2. Impute missing values
3. Feature engineering opportunity: BMI
4. Metrics: Accuracy, Recall

Thank You

github.com/willwu29

Repositories: heart-attack-prediction-model