

---

# Heart Attack Prediction Model

Will Wu

Mar 24<sup>th</sup>, 2025

# Project Overview– Heart Attack Prediction

---

## Problem

- Over 800,000 heart attacks occur annually in the U.S.
- Limitations in Existing Models: 1. Age limitation: 30-79.    2. Require blood test

## Solution

### Early Detection System

- Preventive medications
- Respond quickly



### Machine Learning Models

- Predict risk of heart attack (Yes/No)
- Less limitations

## Potential Impact

- **Even 1%** reduction in heart attack could:
  - Prevent over **8,000** cases annually.
  - Save more then **\$161 million** in health care expenditure.

# Dataset Overview

---

## Sprint 1

---

### Data Source:

- Kaggle
- Originates from CDC 2022 BFRSS Survey Data

### Concerns

- 57% rows contain missing values
- **Less than 250,000** clean observations
- Class imbalance: 5% positive vs. 95% negative



## Sprint 2

---

### Data Source:

- CDC 2022 BFRSS Survey Data
- Add 2023 Survey Data

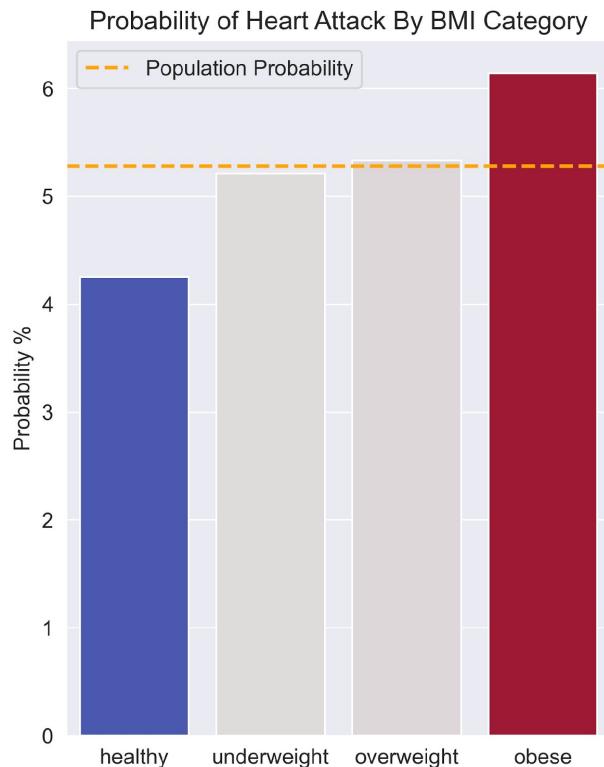
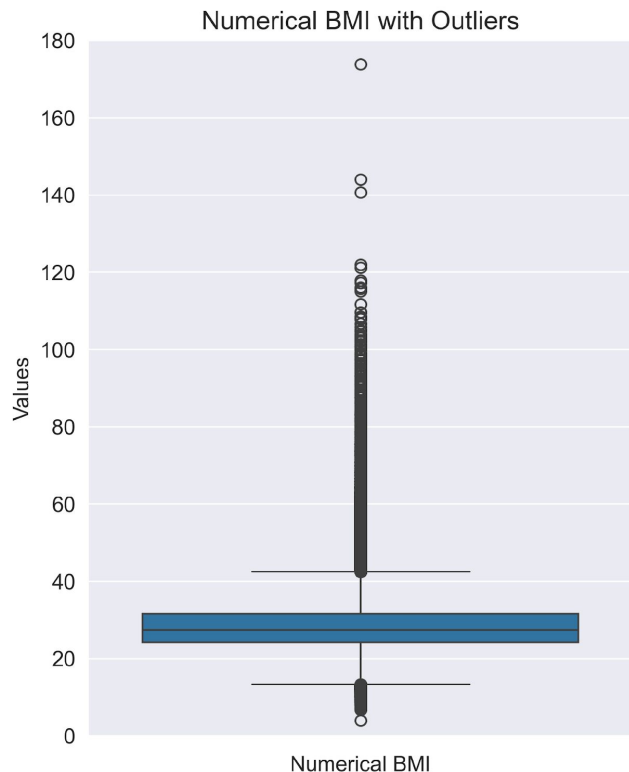
### Results

- **Over 750,000** clean observations.
- 37 features
- 50%-50% balance class in train datasets

### Methods:

- Impute missing values
- Resampling

# EDA Insights and Feature Engineering



## Numerical BMI (Left):

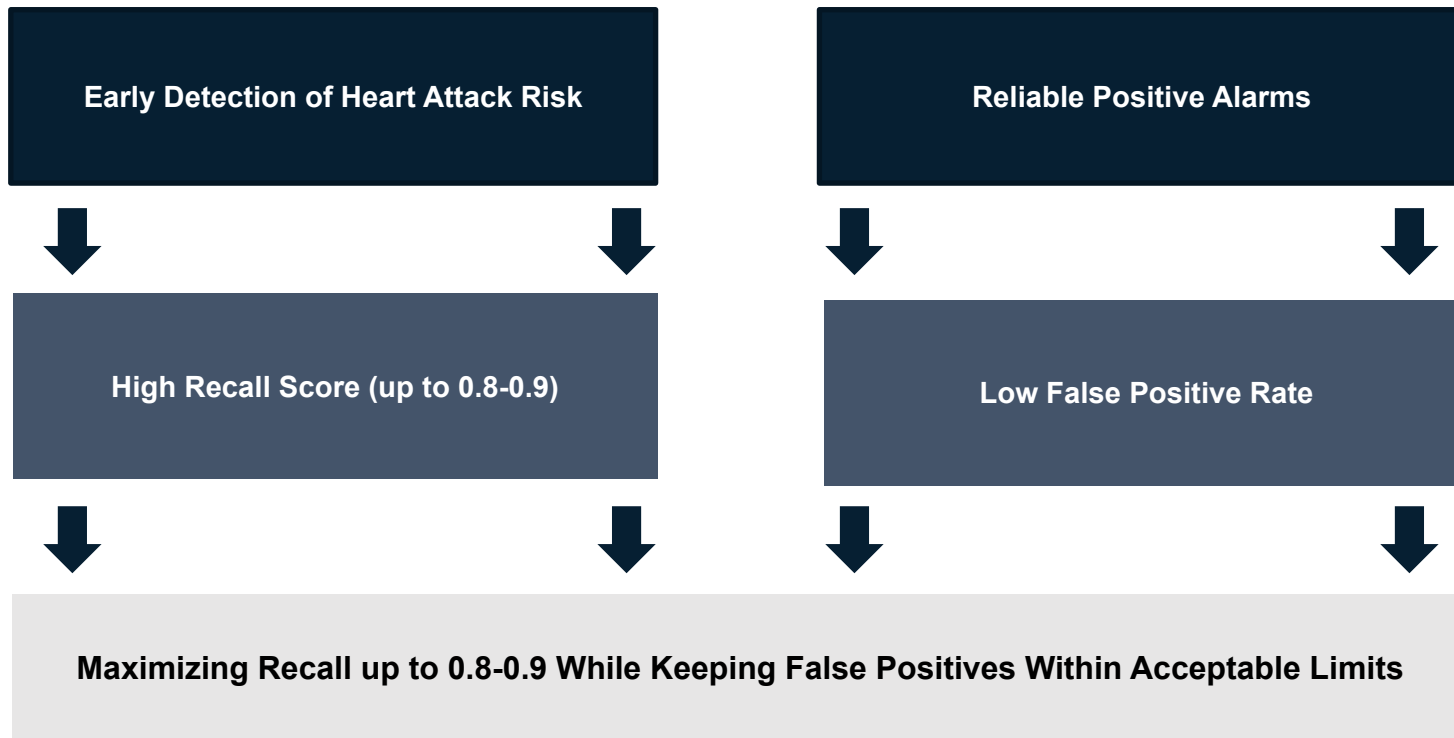
- Outliers

## Bin BMI into categories (Right):

- Remove outliers
- Healthy group: Less likely
- Obese group: More likely

# The Metrics of Success

---



# Baseline Models

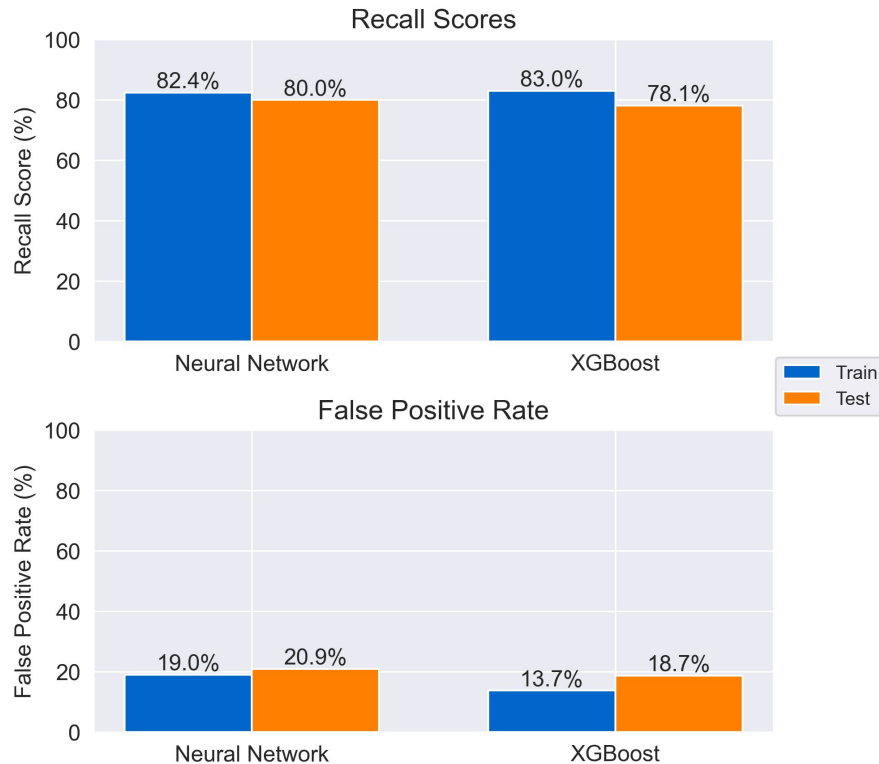
---

## Baseline Models:

- Logistic Regression
- Naïve Bayes
- Decision Tree
- Random Forest
- XGBoost
- Neural Network

## Top performing models:

- Neural network slightly outperformed in recall
- XGBoost exhibits signs of overfitting

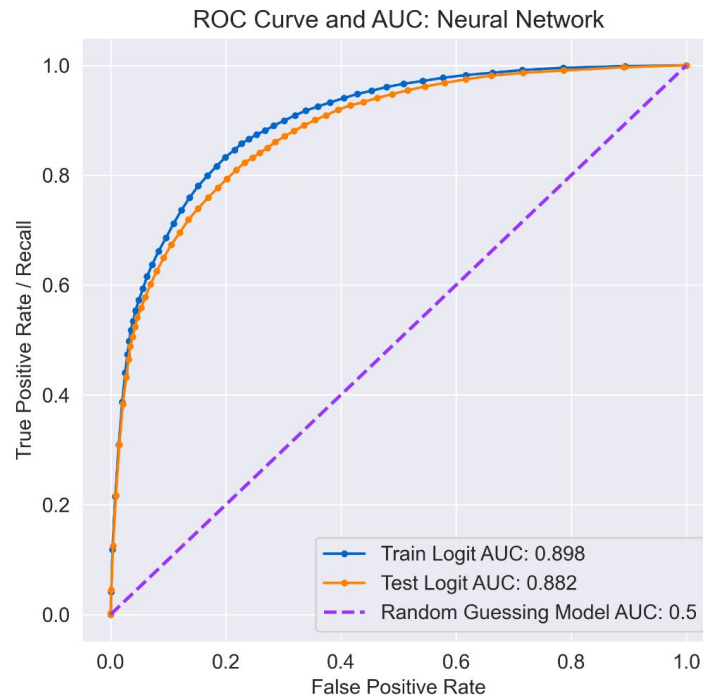
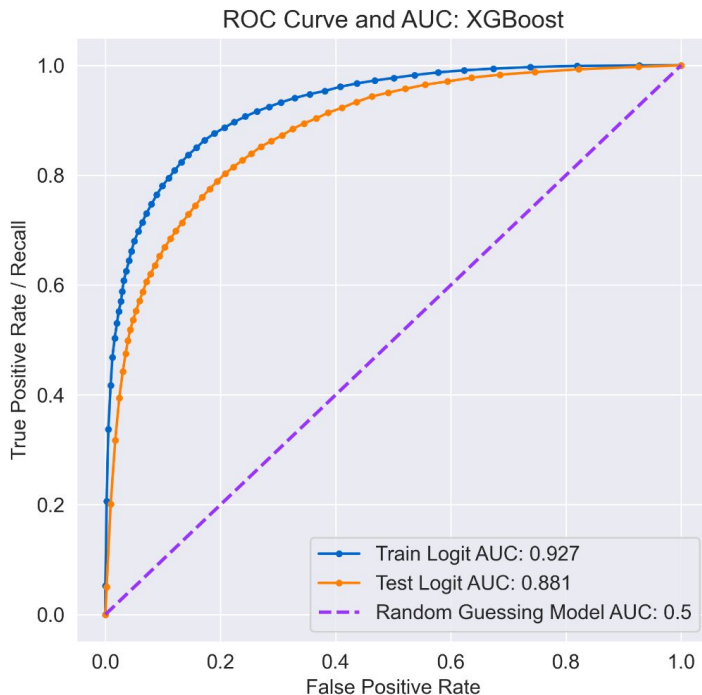


# XGBoost vs. Neural Network

Relatively High AUC

## Overfitting:

- Higher variance
- More likely to underperform



---

## Next steps

- 1 Feature selection
- 2 Customize loss function for neural network
- 3 Advanced modeling



---

# Thank You

[github.com/willwu29](https://github.com/willwu29)

Repositories: heart-attack-prediction-model