

# **HEART ATTACK INDICATORS**



**HAI**

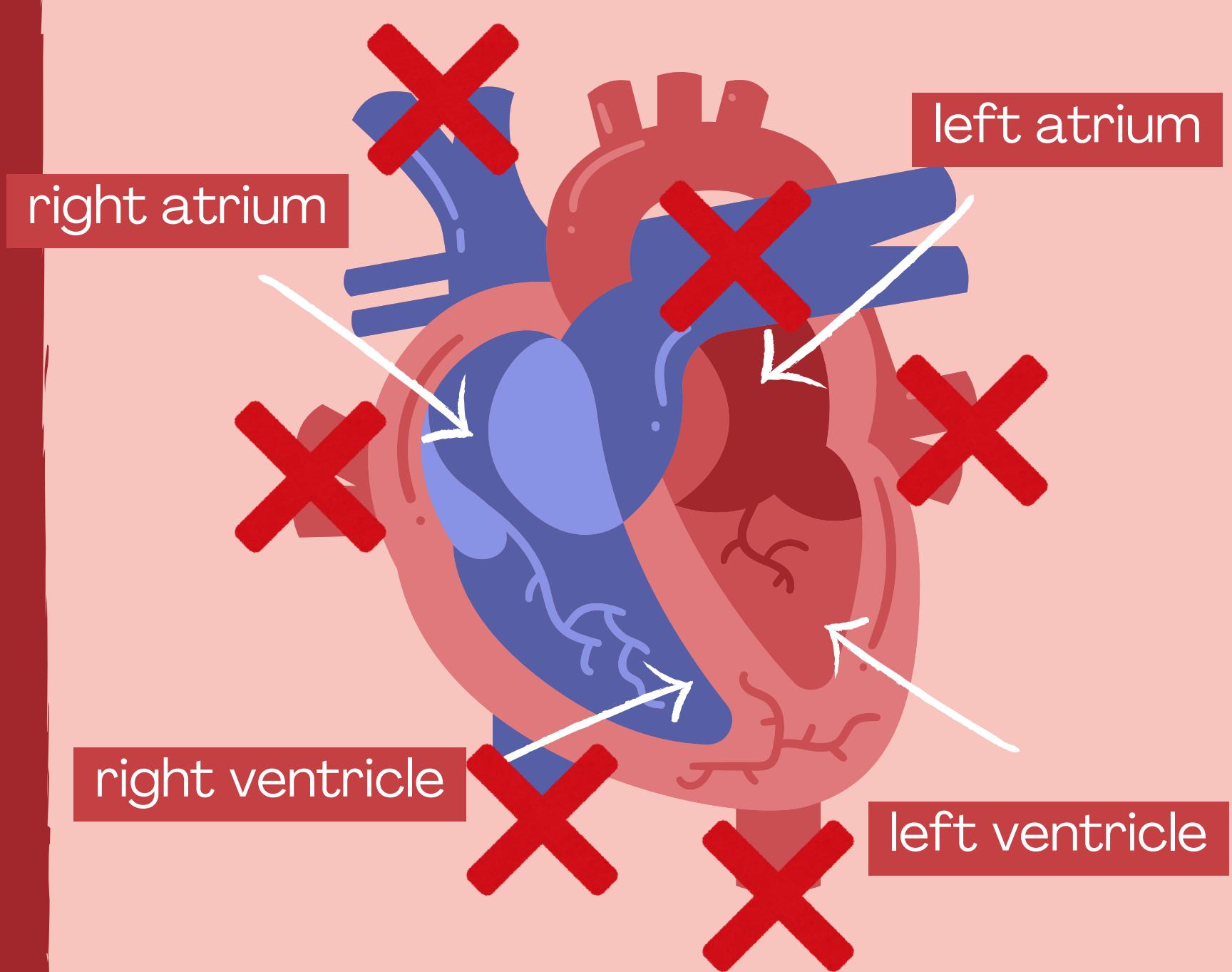
By  
**Stefan George**

# HERES THE PROBLEM

Heart Disease, is a blanket term for a few different conditions all can lead to a heart attack.

What if we could detect the symptoms sooner

How many lives could be saved



# DATA OVERVIEW

---



Data

CDC

Format

Survey

Answers

Self  
Reported

Specifics

40  
+440k

# DATA OVERVIEW CONTINUED



Cleaning



Pre processing



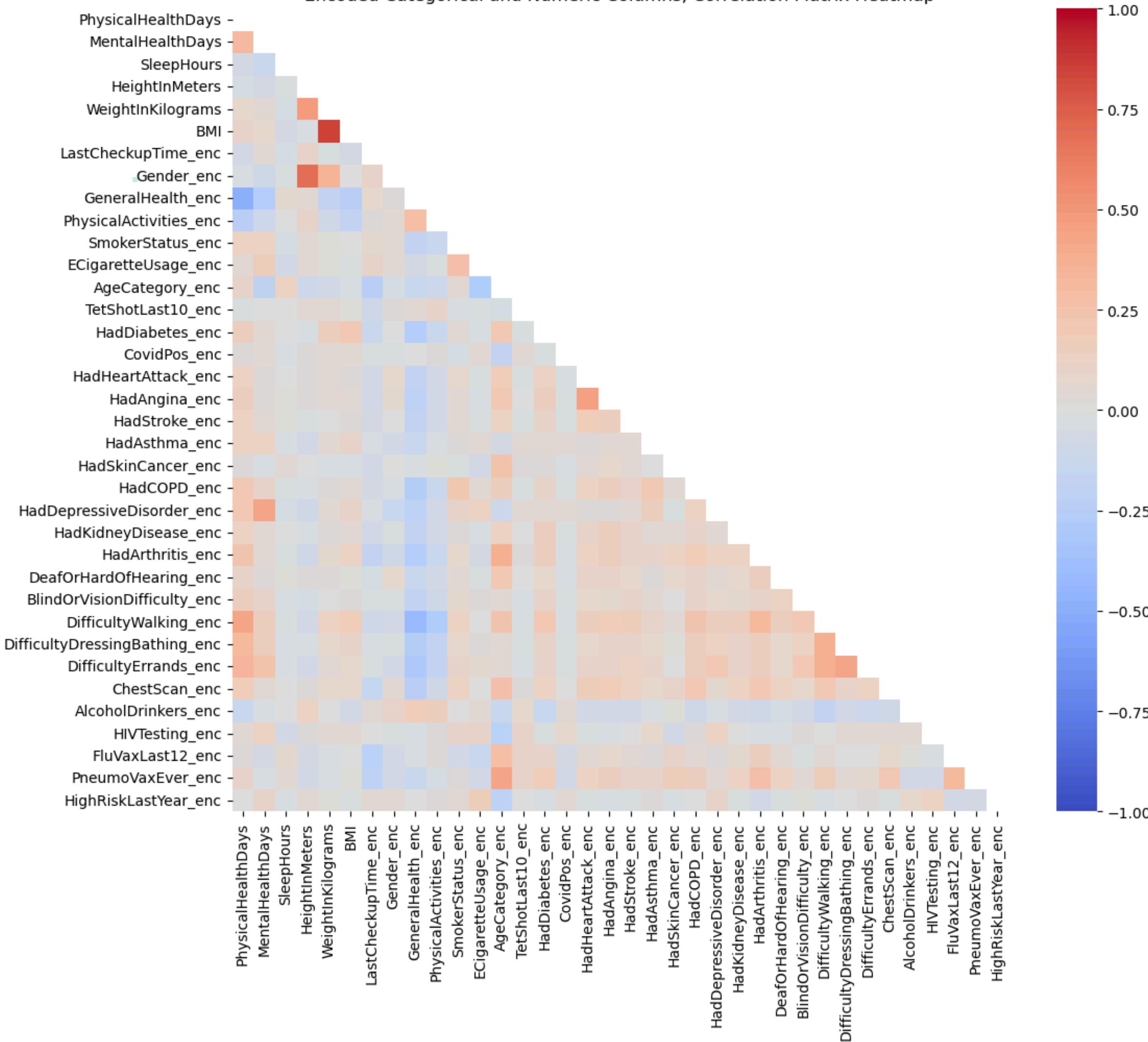
Encoding

Data

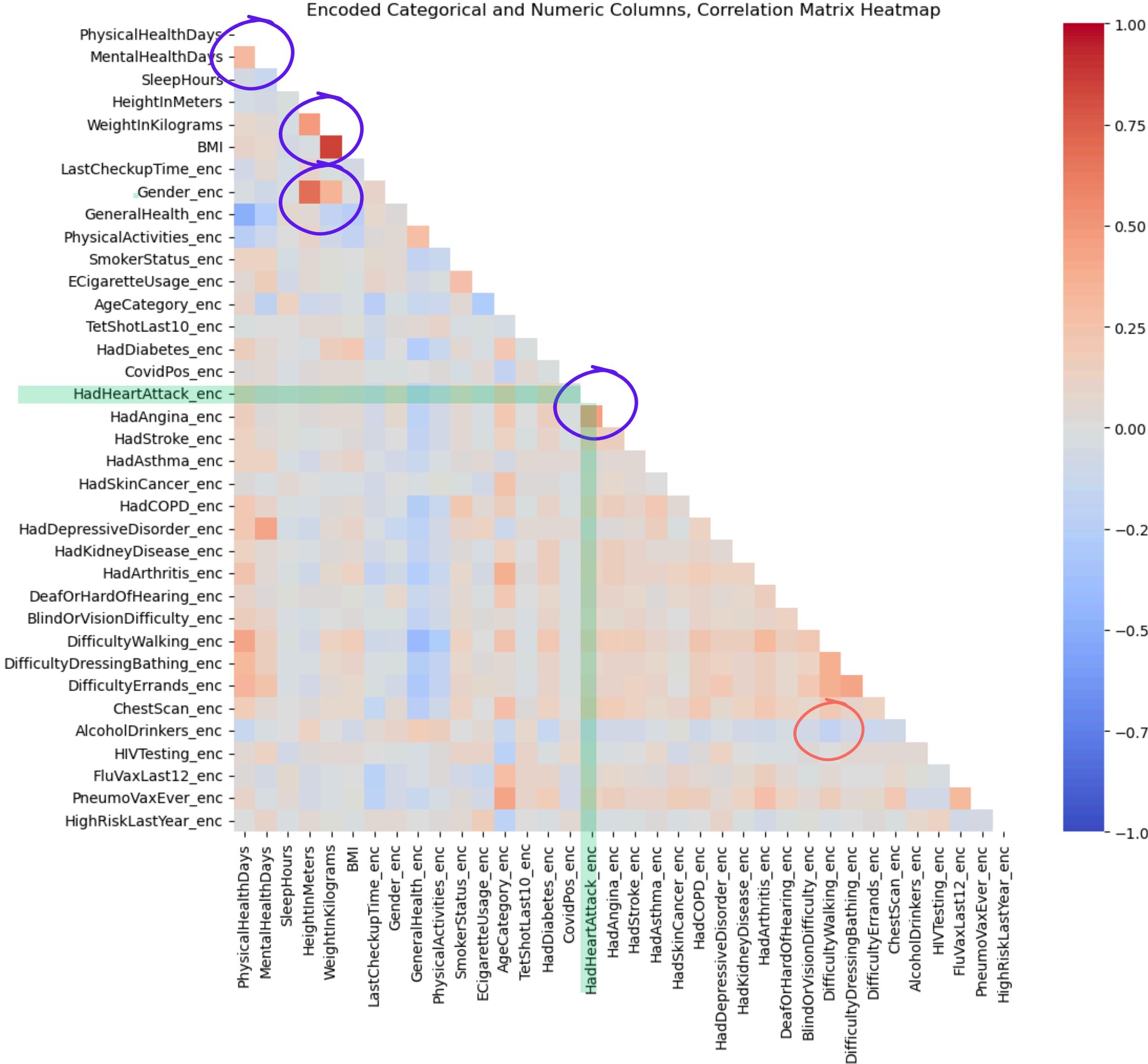
36  
+240k

# THE DATA...

Encoded Categorical and Numeric Columns, Correlation Matrix Heatmap



# THE DATA SPEAKS



# SOLUTION



To make a early warning / early detection system for people who are at greater risk of heart attacks.

# MODEL METRICS

Key Metrics

Recall

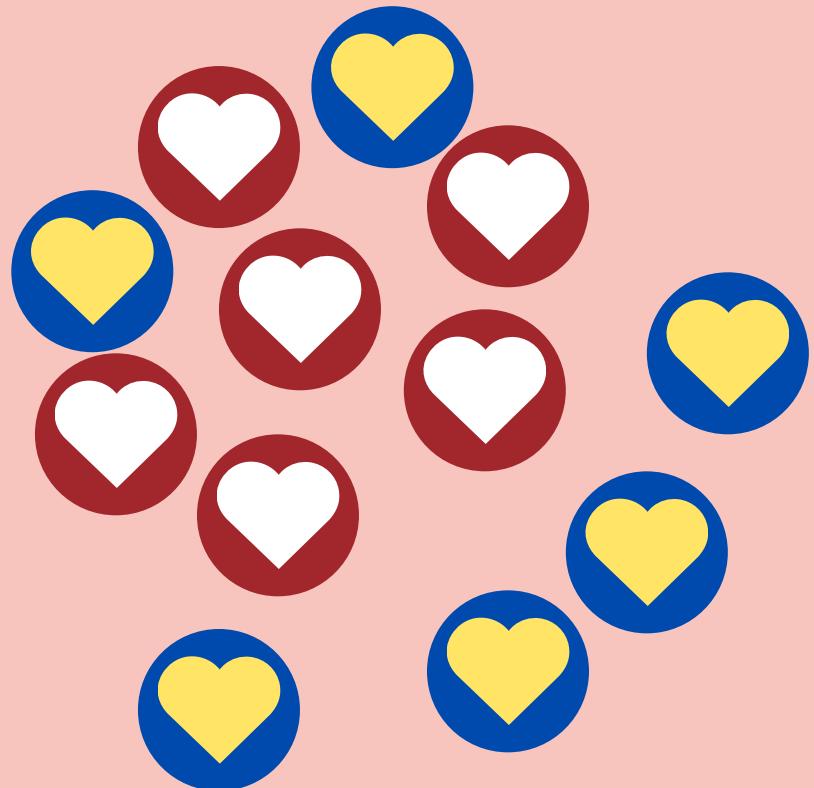
Precision



# MODEL METRICS

## Key Metrics

Recall

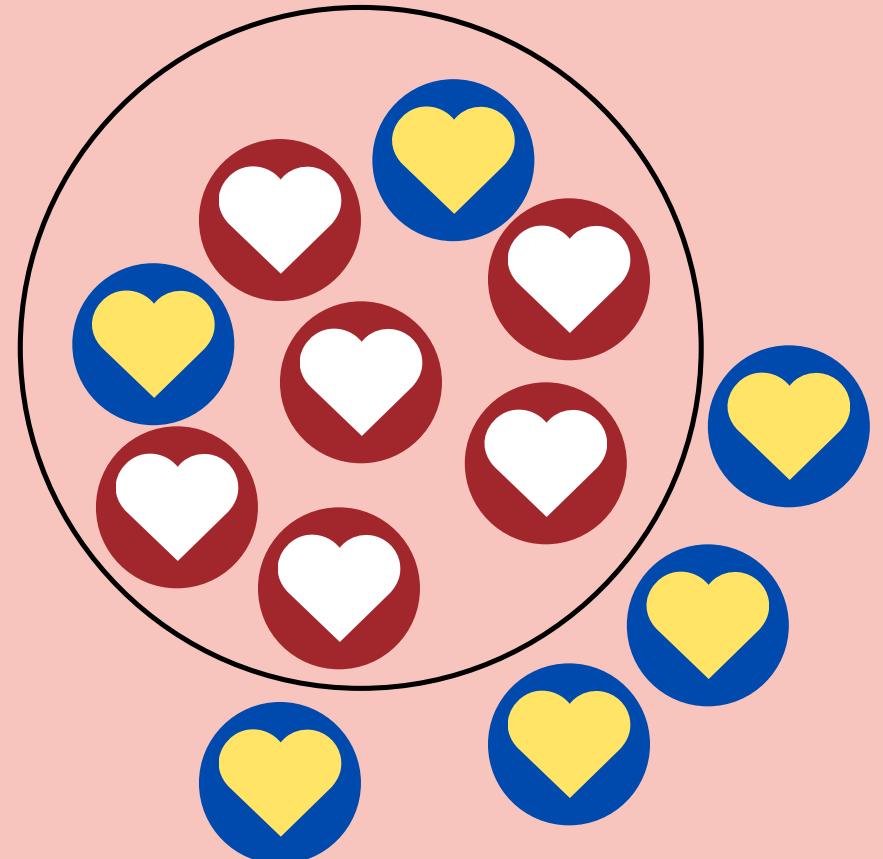


Precision

# MODEL METRICS

## Key Metrics

Recall

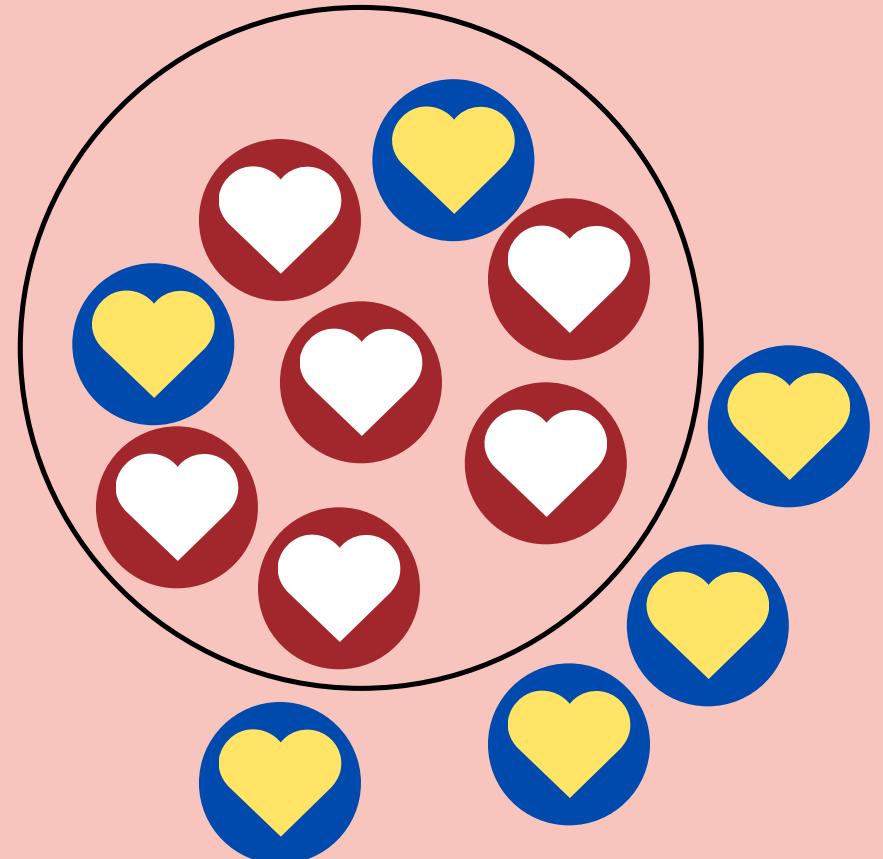


Precision

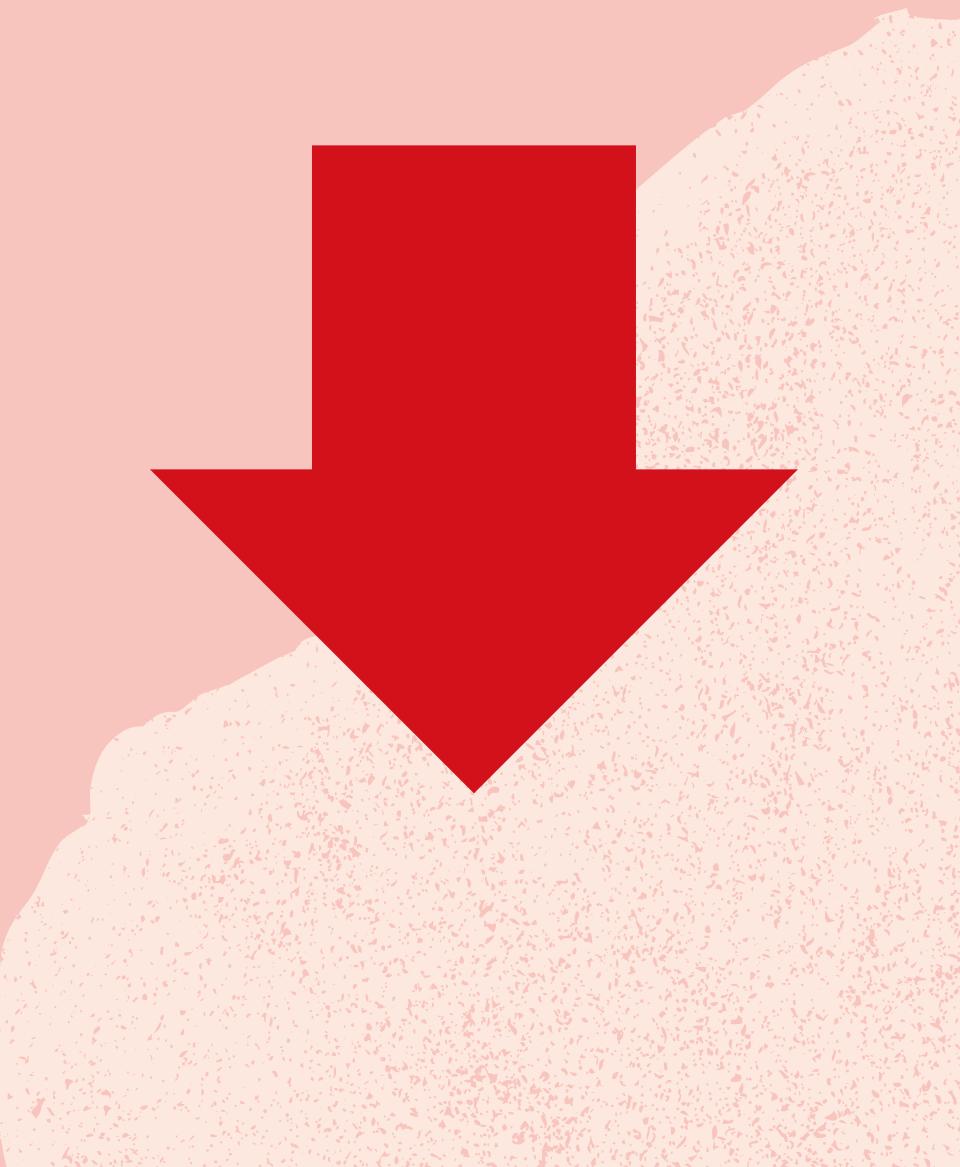
# MODEL METRICS

## Key Metrics

Recall



Precision



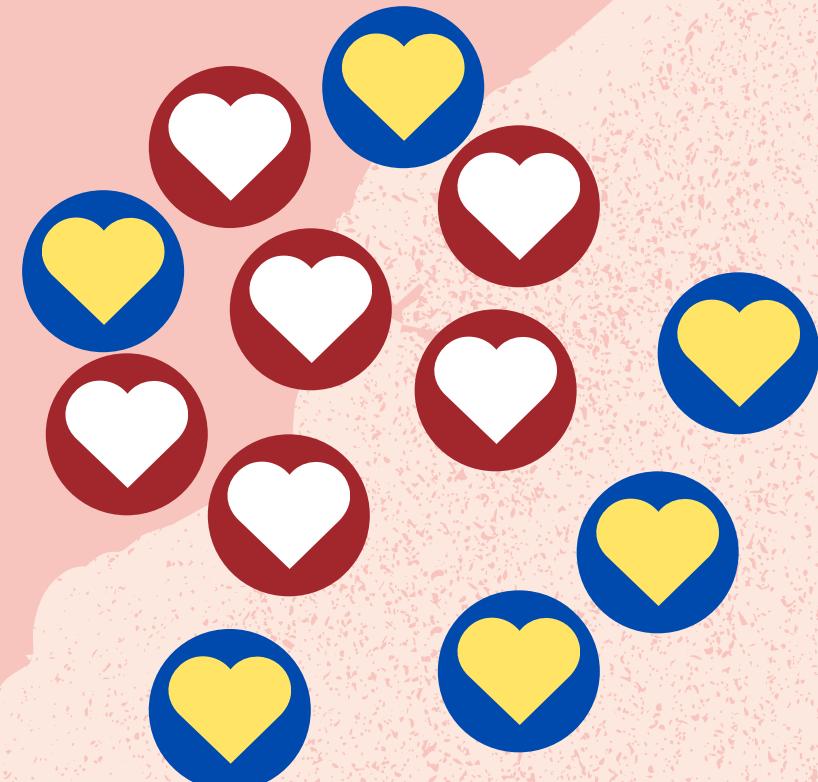
# MODEL METRICS

## Key Metrics

Recall



Precision



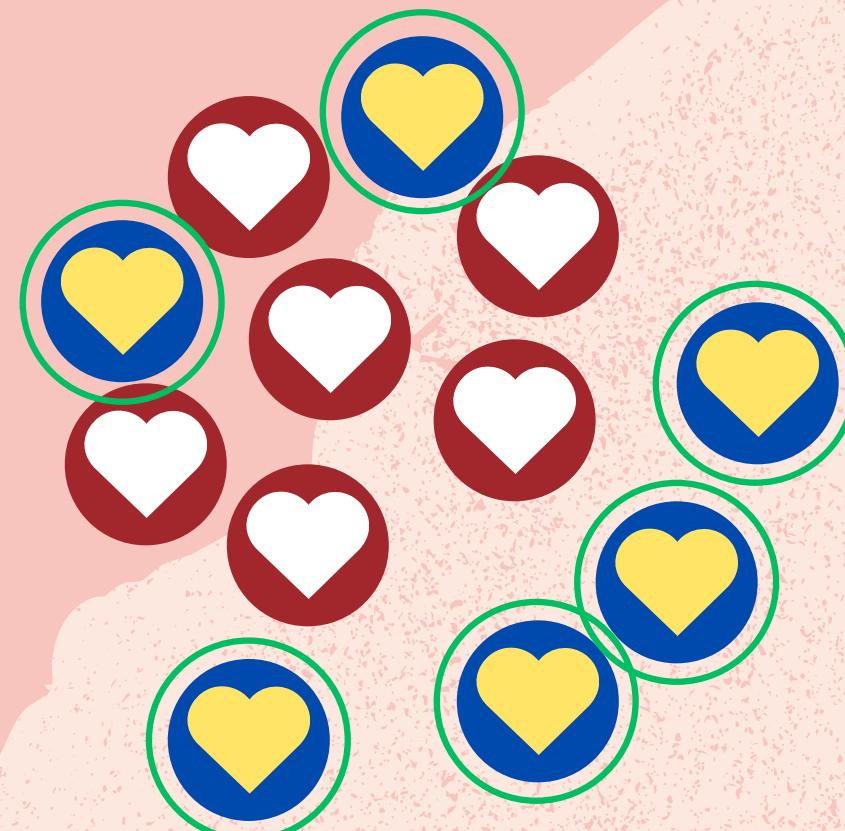
# MODEL METRICS

## Key Metrics

Recall



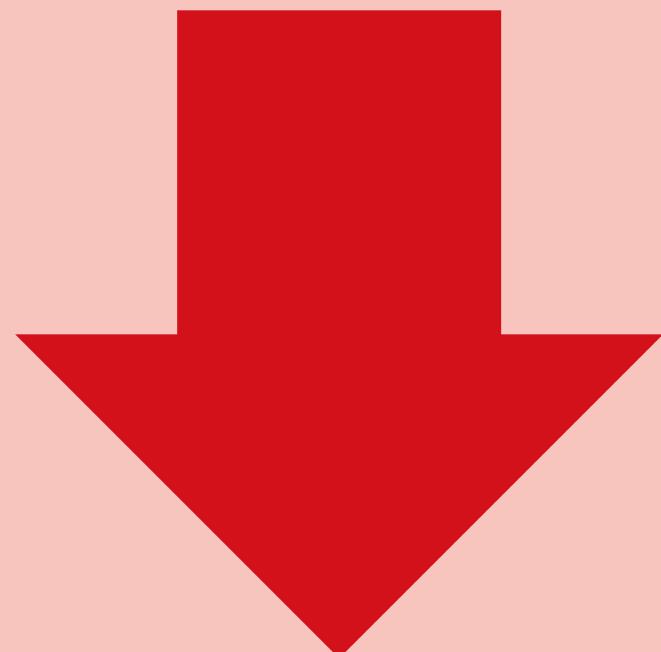
Precision



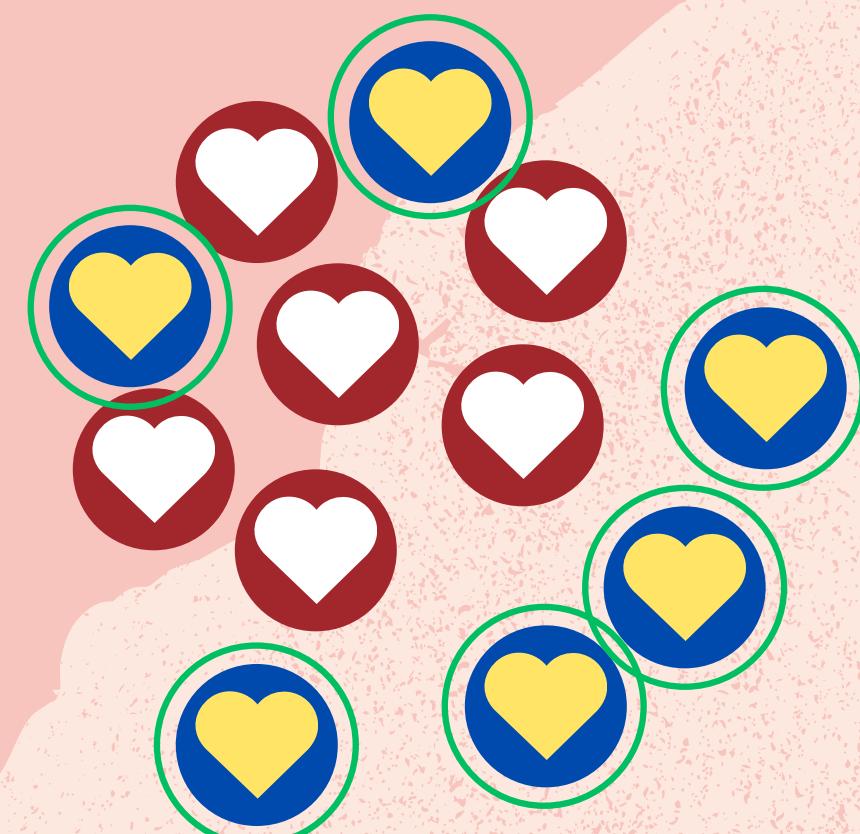
# MODEL METRICS

## Key Metrics

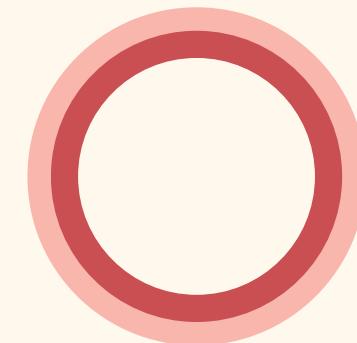
Recall



Precision



# MY SECRET TECHNIQUES



## Models

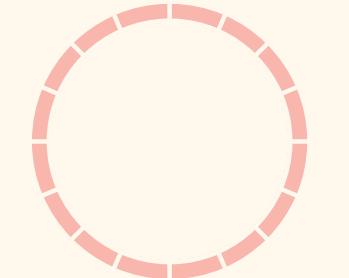
Logistic Regression

Decision Tree

Random Forest

KNN

XGBoost



## Techniques

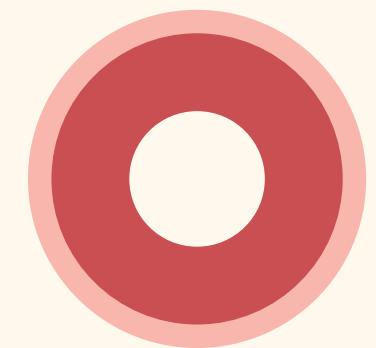
Baseline Model

Optimized Models  
using Hyperparameters

Threshold Boosted

SMOTE

Feature Removal



## Best Final Model

# BASELINE TEST PERFORMANCE EVALUATION

Baseline Logistic  
Regression

**Recall = 24%**

PRECISION - 55%

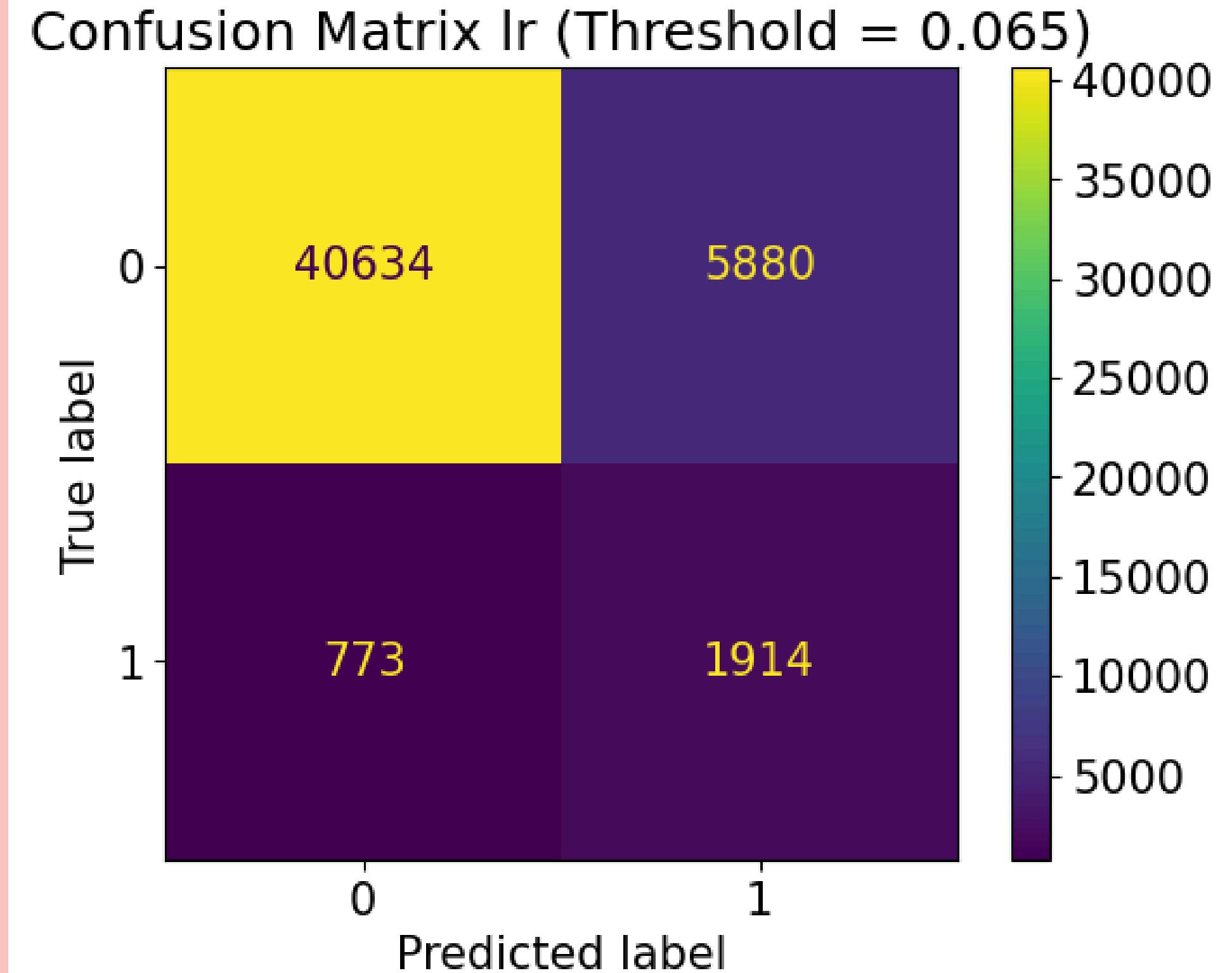
# TEST PERFORMANCE EVALUATION

Threshold  
0.065

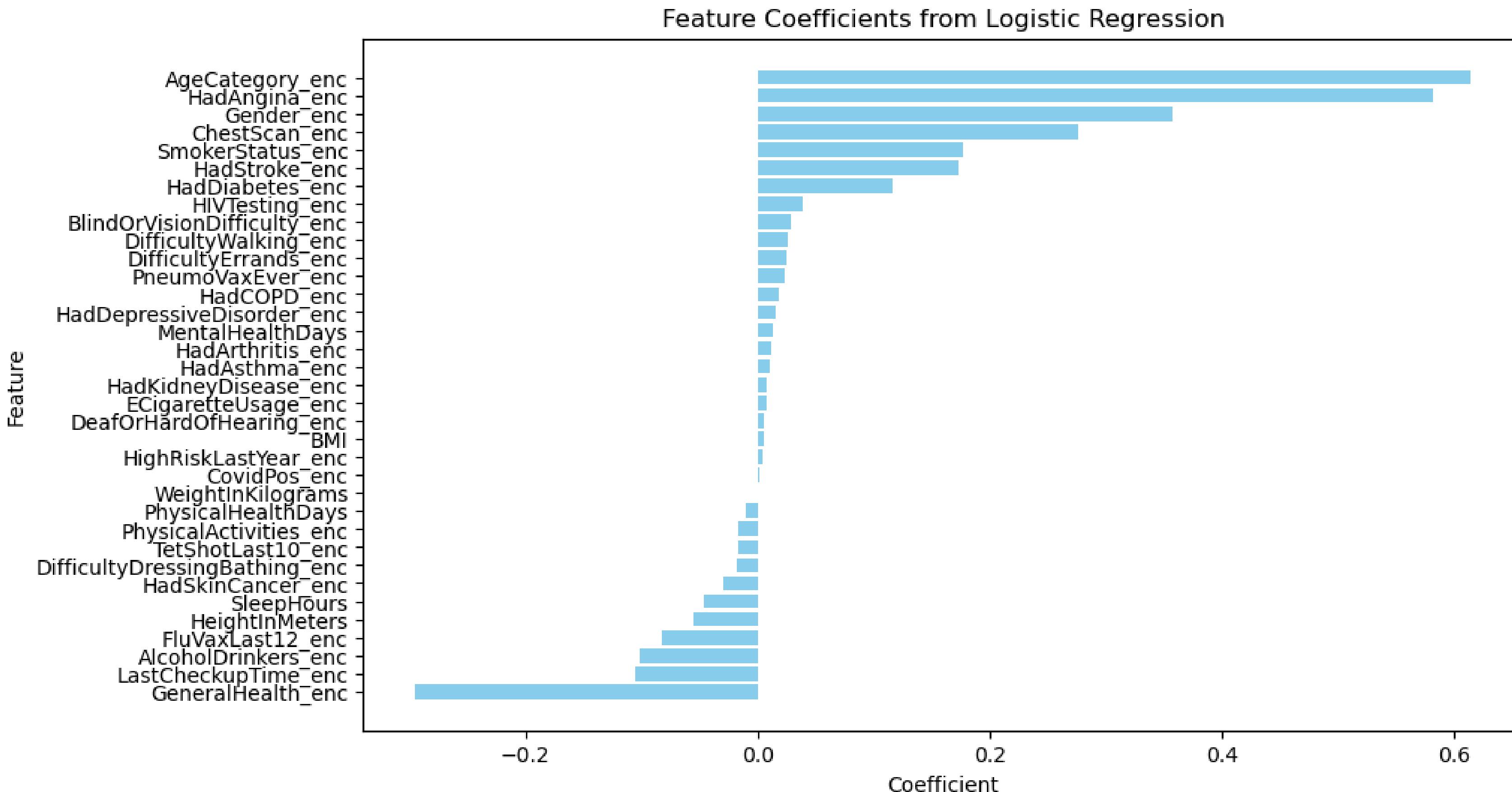
**Best Logistic  
Regression**

**Recall = 71%**

**PRECISION - 25%**

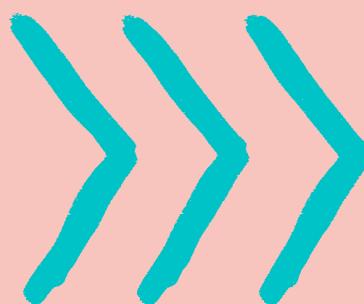


# BEST LR FEATURE COEFFICIENTS



# TEST PERFORMANCE EVALUATION

Threshold



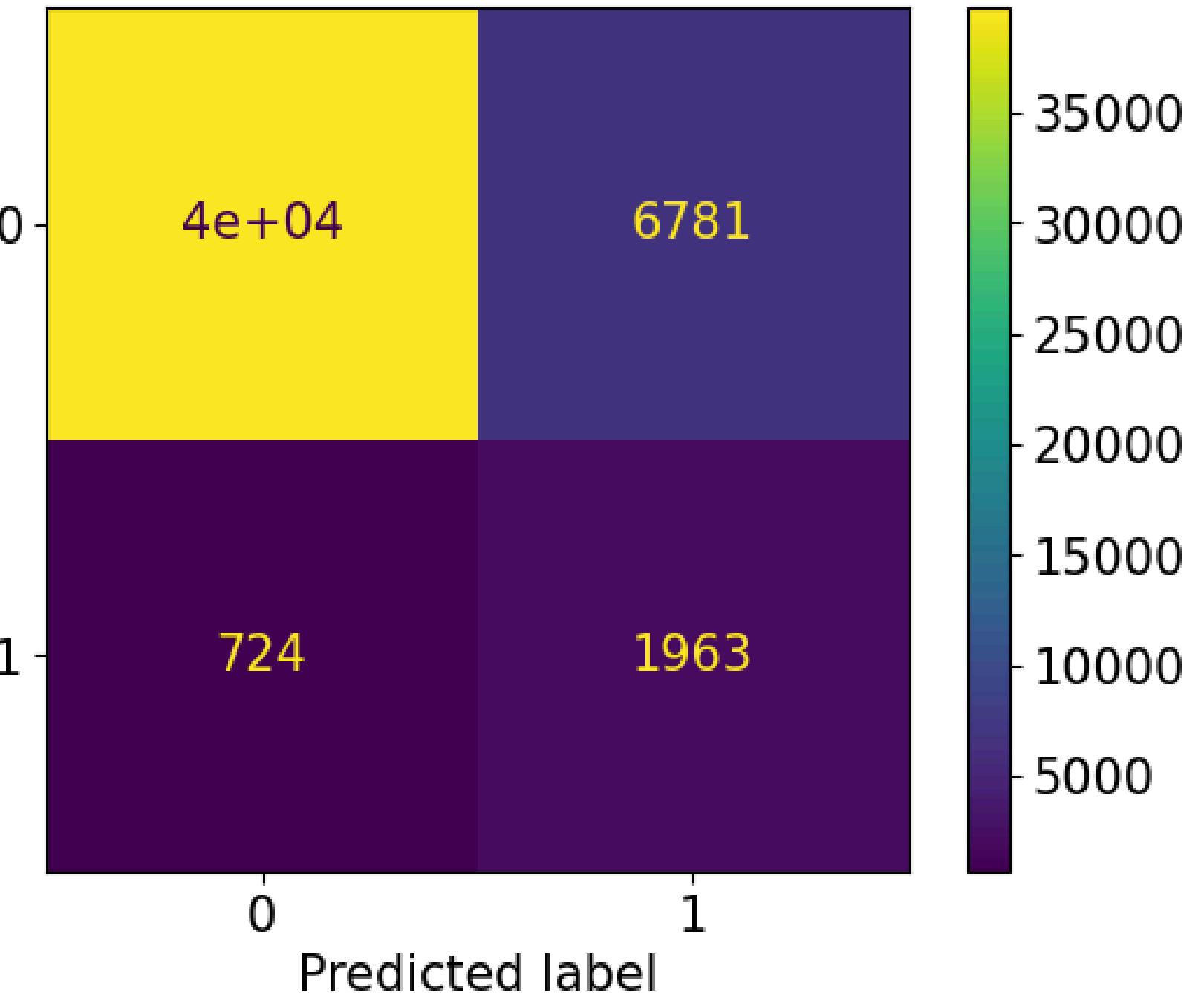
0.065

**Best XGBoost**

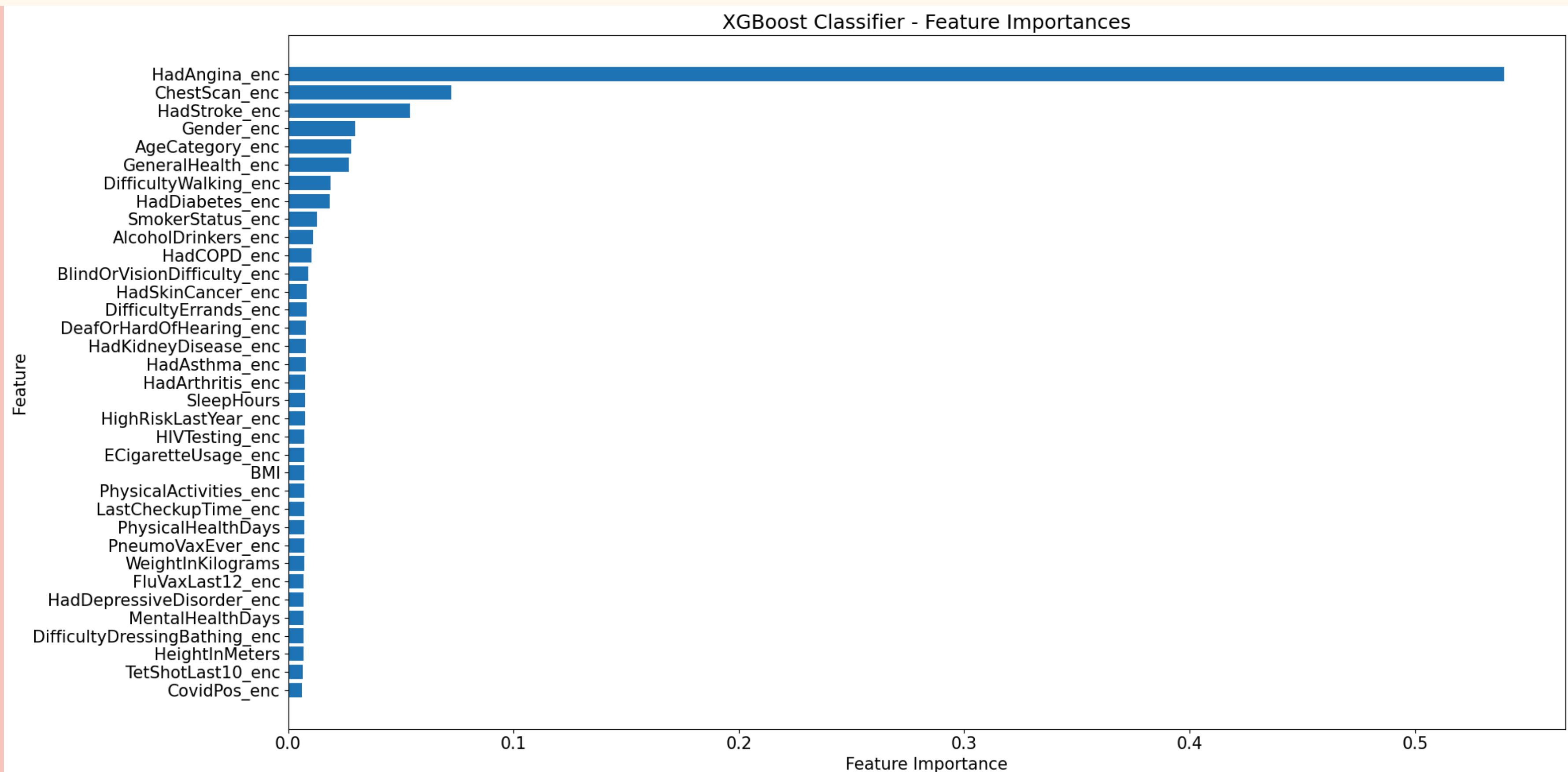
**Recall = 73%**

**PRECISION - 23%**

Confusion Matrix XGBoost (Threshold = 0.065)



# BEST XGBOOST FEATURE COEFFICIENTS



# TEST PERFORMANCE EVALUATION

## Baseline Logistic Regression

**Recall = 24%**

PRECISION - 55%

## Best Logistic Regression

**Recall = 71%**

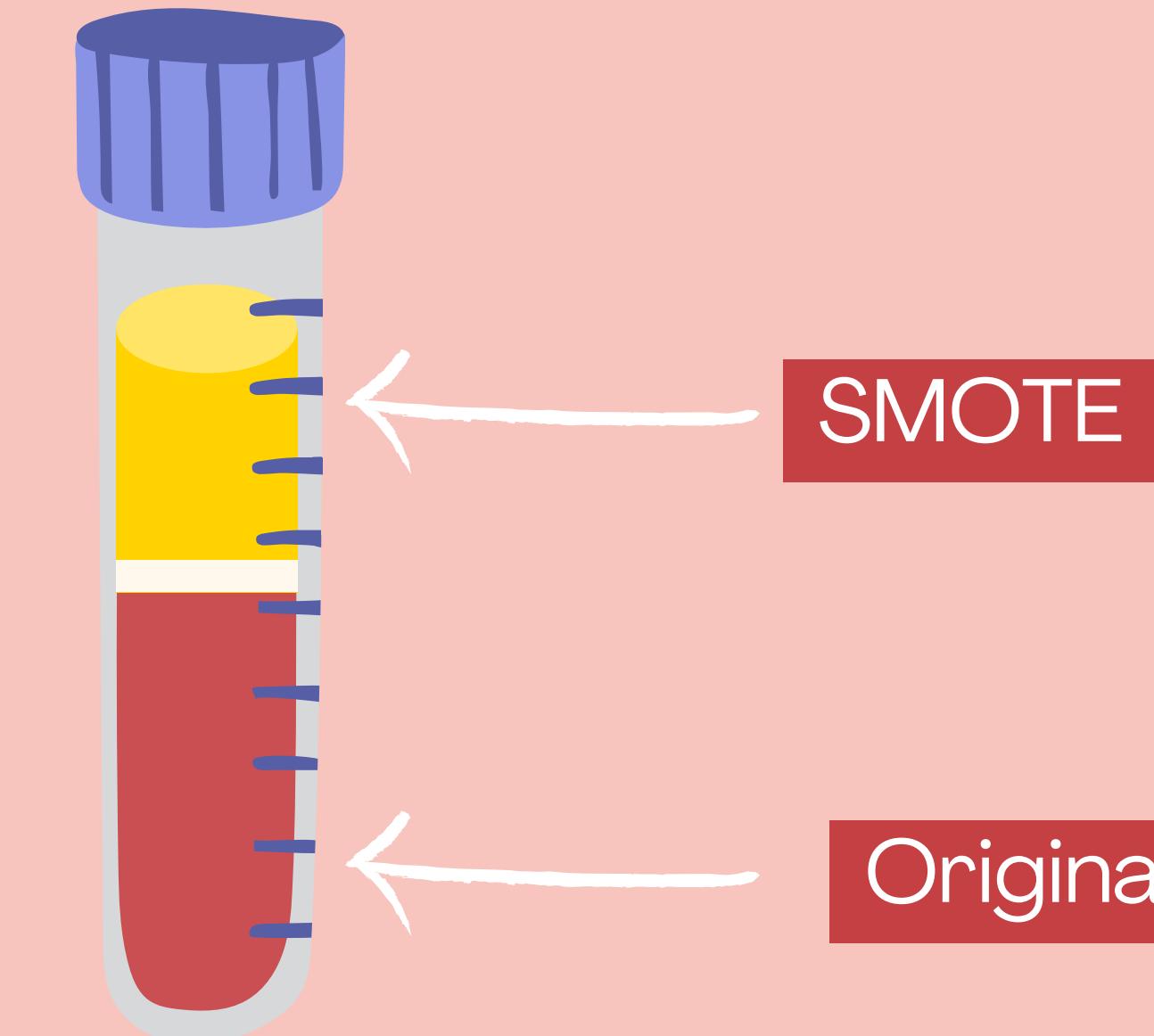
PRECISION - 25%

## Best XGBoost

**Recall = 73%**

PRECISION - 23%

# SMOTE



Smote adds synthetic data to an imbalanced feature.

HadHeartAttack

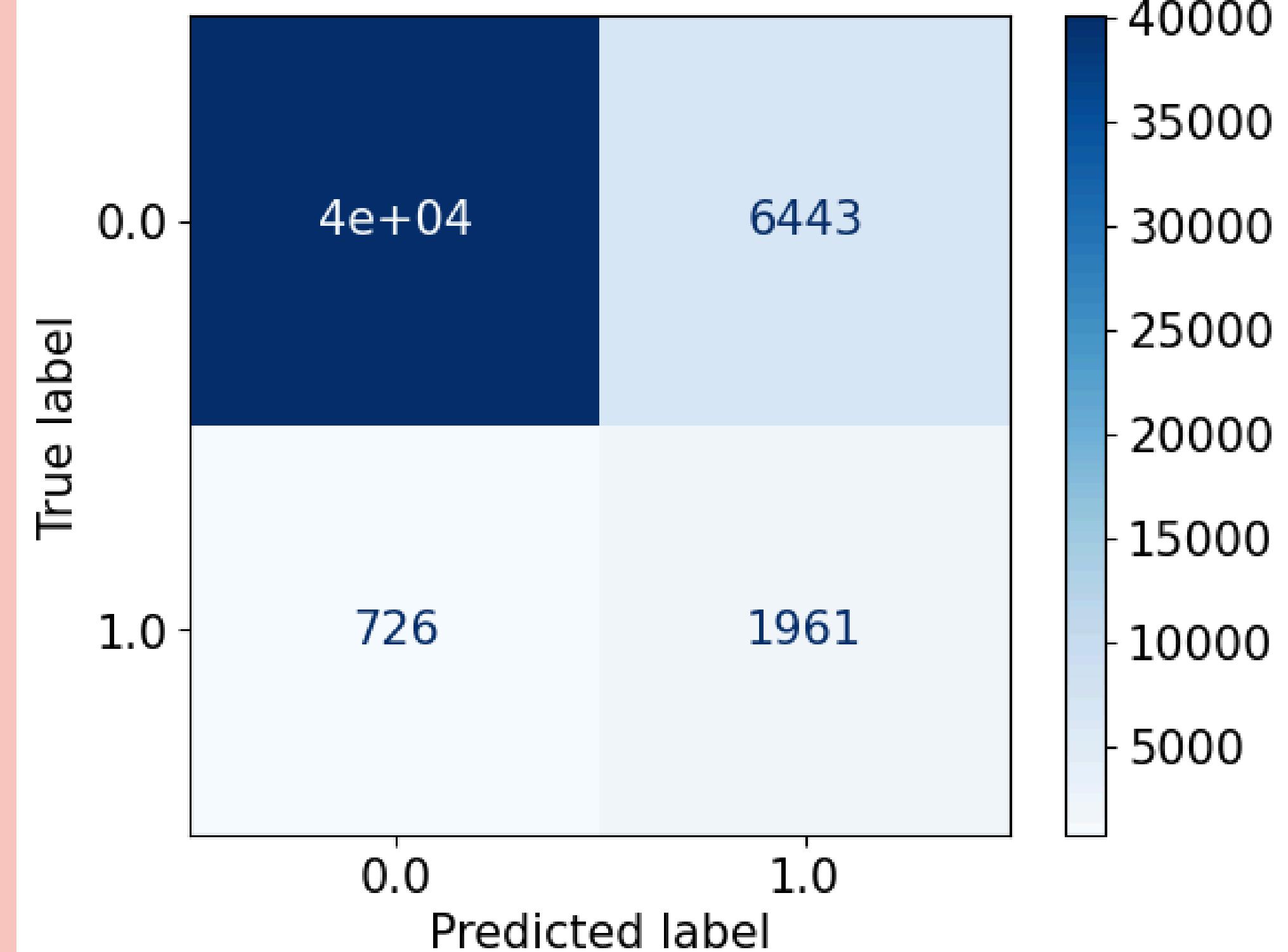
# SMOTE LR TEST PERFORMANCE EVALUATION

## Smote Logistic Regression

**Recall = 73%**

**PRECISION - 23%**

Confusion Matrix - Test



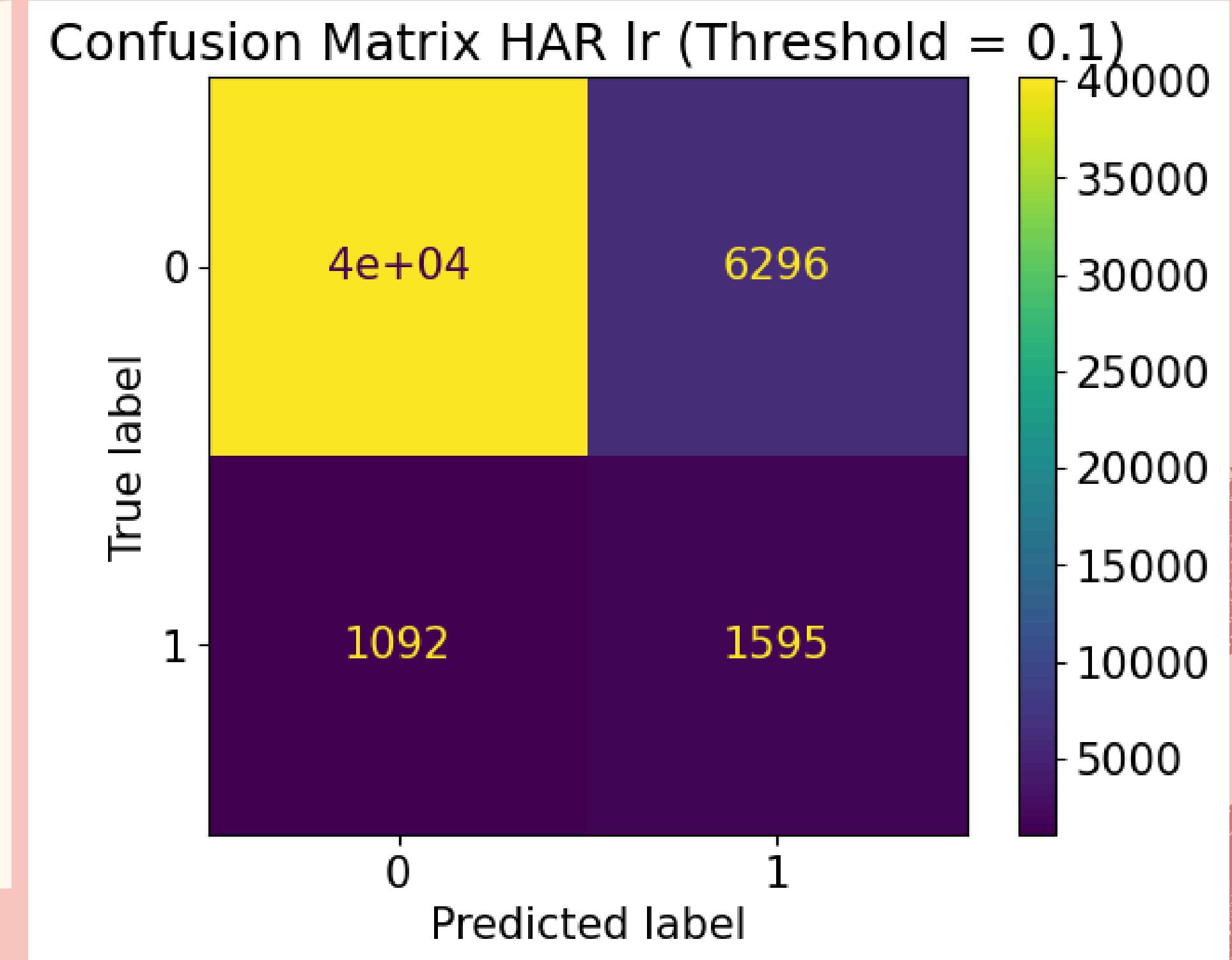
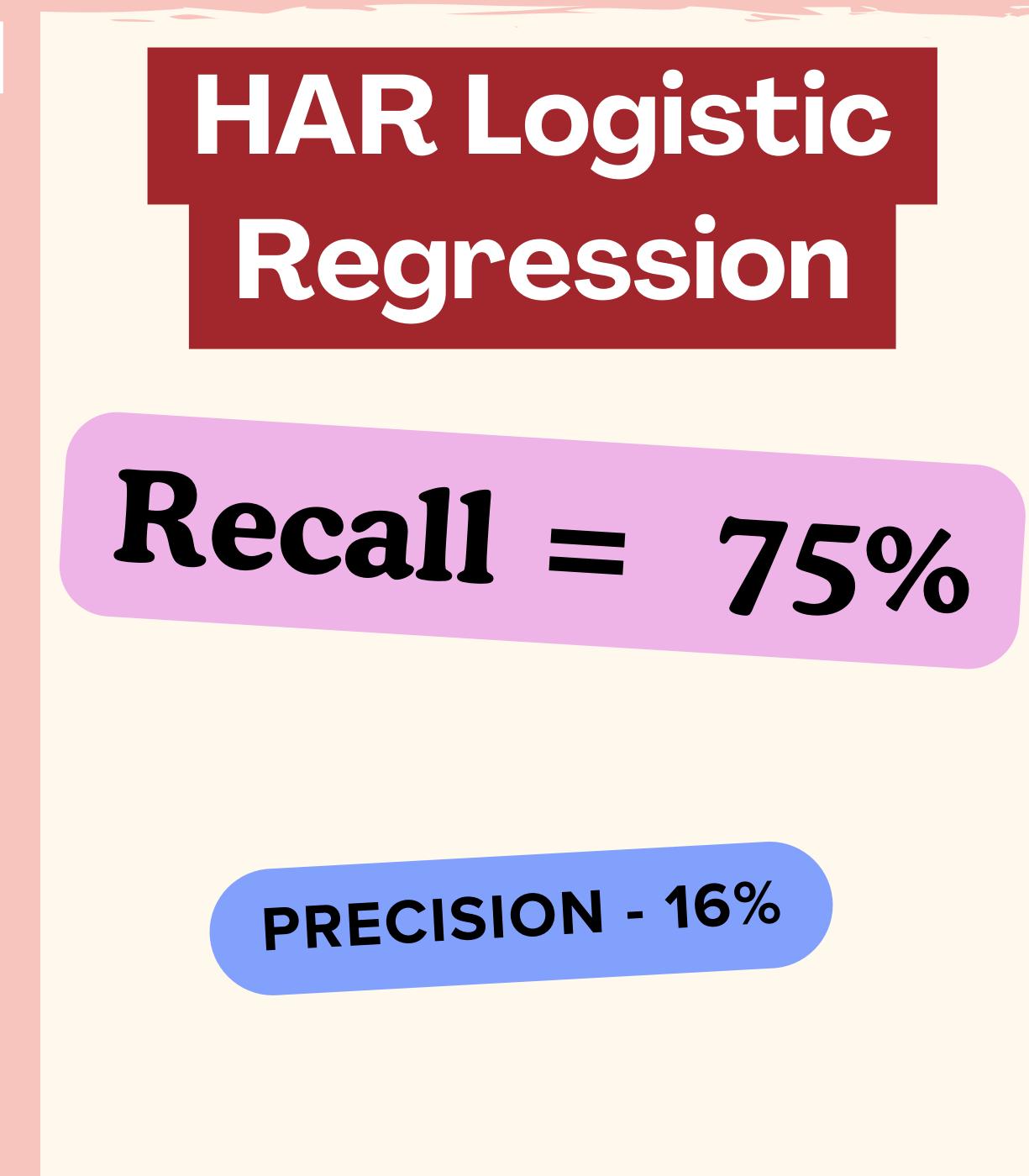


# HAD ANGINA

Data  
HadAngina

# HADANGINA REMOVED LR TEST PERFORMANCE EVALUATION

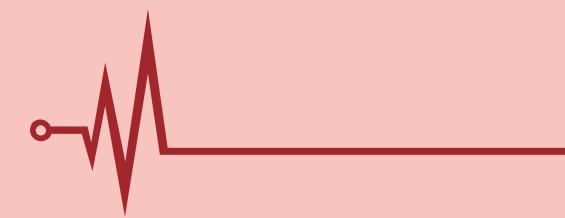
Threshold  
0.065



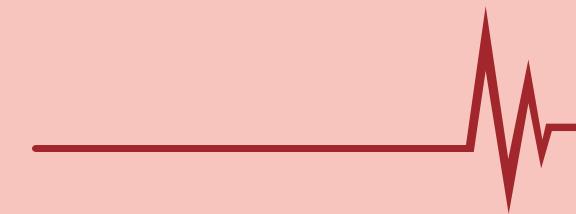
# MODEL RANKING

Rank	Models	Recall
1	XGBoost	73%
2	Smote Logistic Regression	73%
3	Best Logistic Regression	71%
4	HAR Logistic Regression	75%
Base	Baseline	24%

# HAI PROJECT IMPACT



Early Diagnosis



Cost



Wait Times

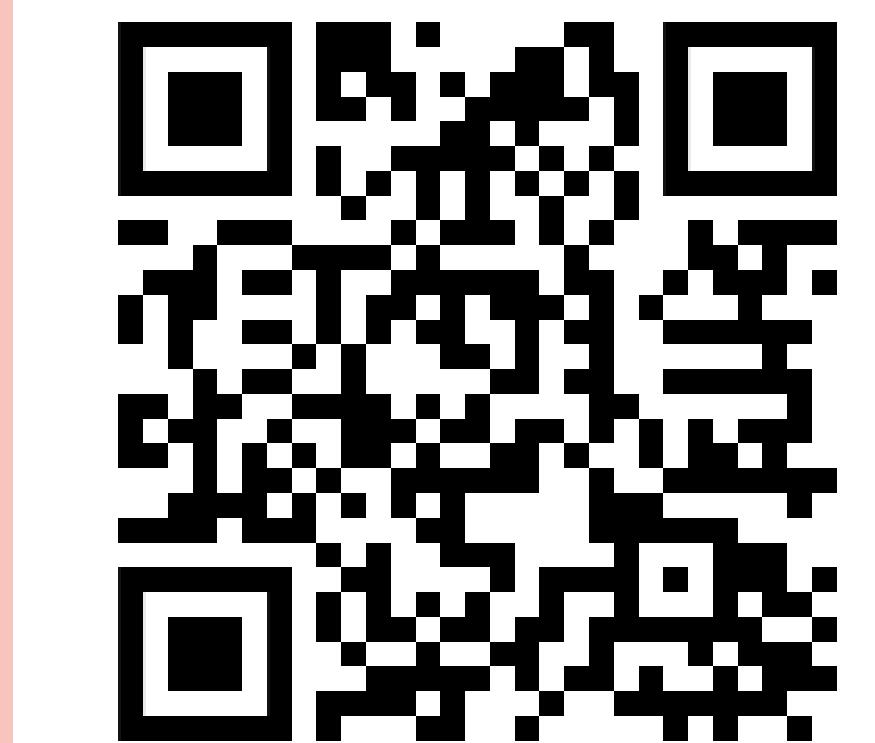


**THANK YOU  
QUESTIONS?**

**LinkedIn**



**GitHub**



# PRESS THESE KEYS WHILE ON PRESENT MODE!

**B** for blur

**C** for confetti

**D** for a drumroll

**M** for mic drop

**O** for bubbles

**O** for quiet

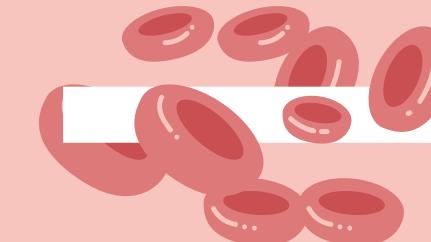
**U** for unveil

**0-9** Any number from  
0-9 for a timer

# **APPENDIX**

# PRE-PROCESSING

**TARGET**



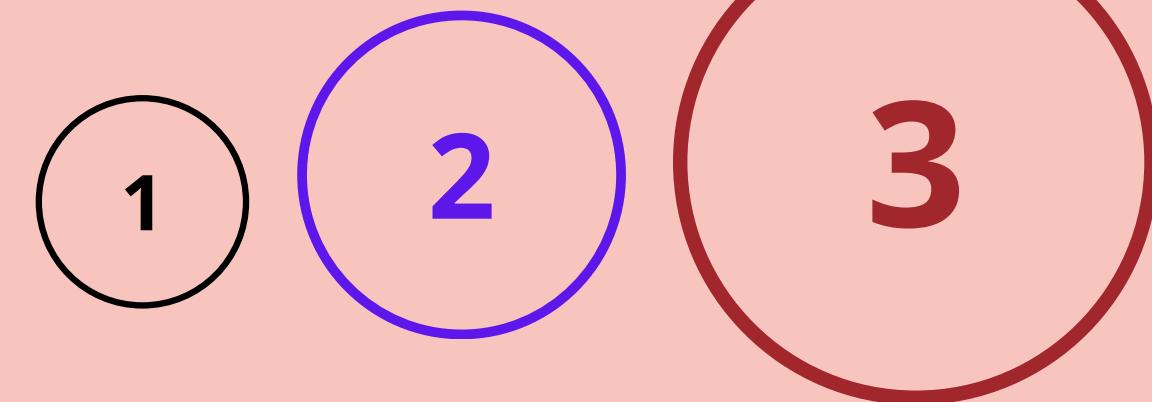
**HADHEARTATTACK**

**Binary**

101010
101010
101010

**ENCODING**

**Ordinal**



- Gender (Sex)
- Yes and No

- Order of Severity
- General Health
- SmokerStatus
- ECigaretteUse