

# **HEART ATTACK INDICATORS**

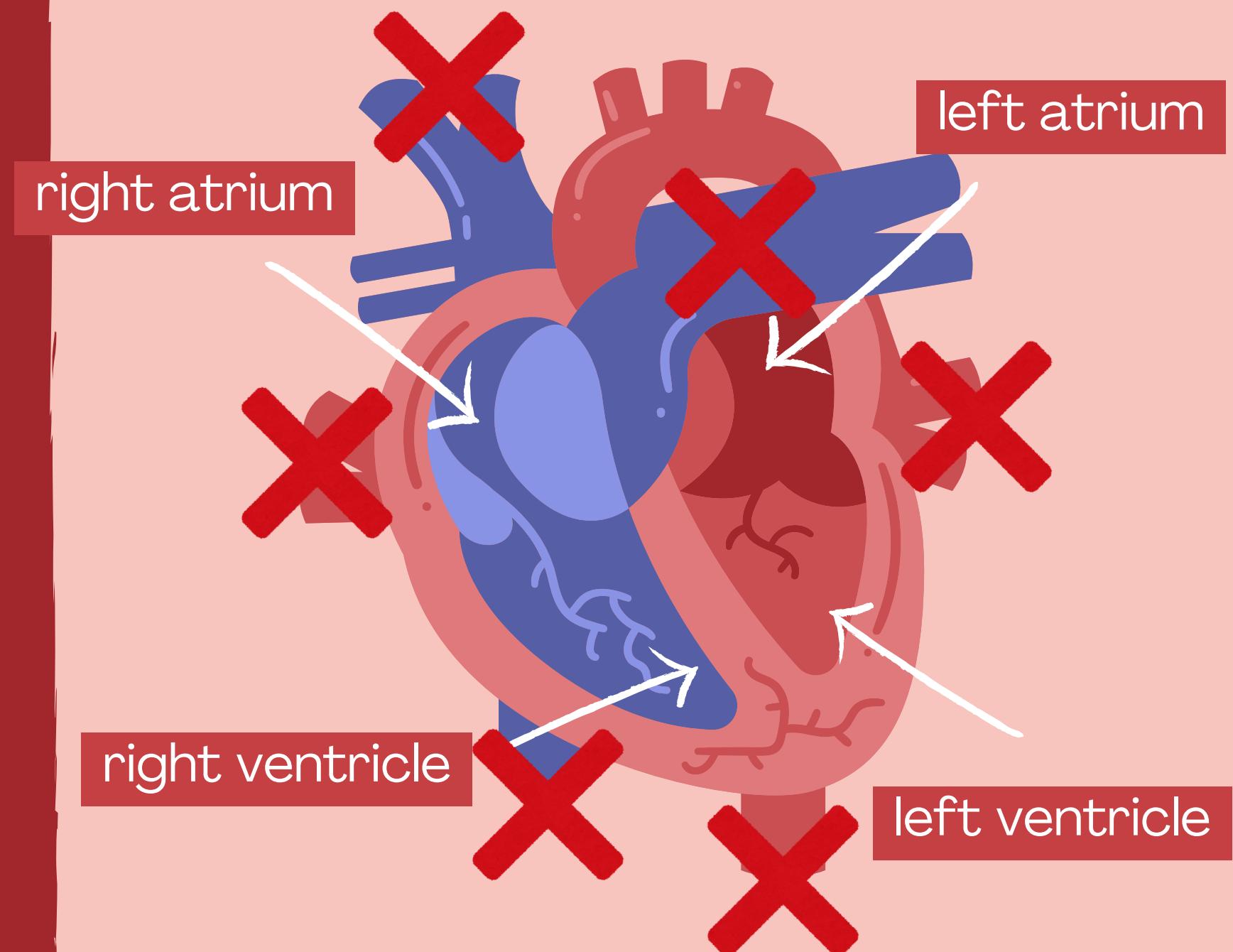


**HAI**

**Capstone Sprint 3  
Stefan George**

# HERES THE PROBLEM

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



# DATA OVERVIEW

---



Data



Data

Solution



Key metric

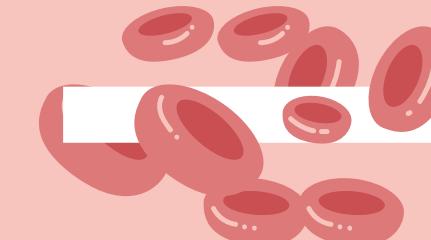


Whats next?



# PRE-PROCESSING

**TARGET**



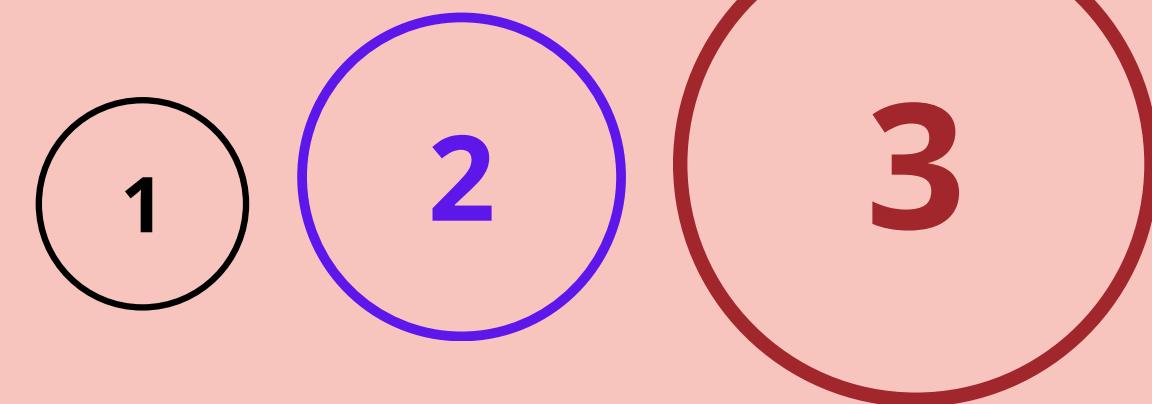
**HADHEARTATTACK**

**Binary**

101010
101010
101010

**ENCODING**

**Ordinal**

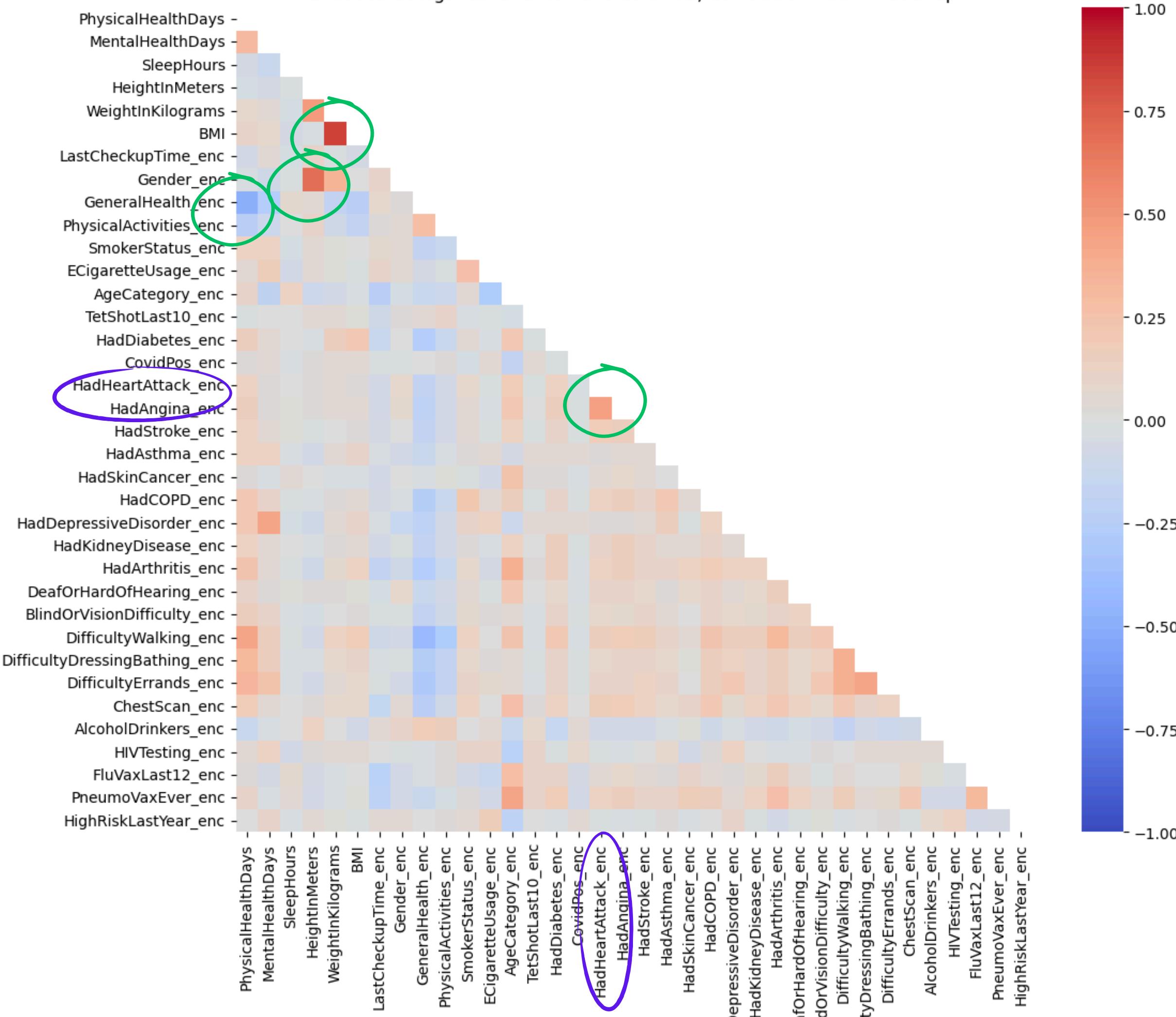


- Gender (Sex)
- Yes and No

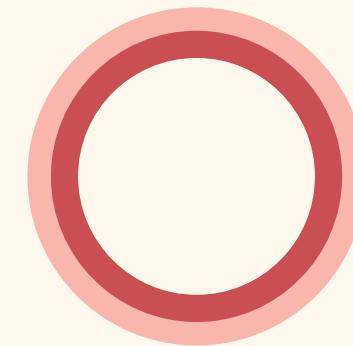
- Order of Severity
- General Health
- SmokerStatus
- ECigaretteUse

# EDA

Encoded Categorical and Numeric Columns, Correlation Matrix Heatmap



# MODELS & HYPERPARAMETERS



## Models

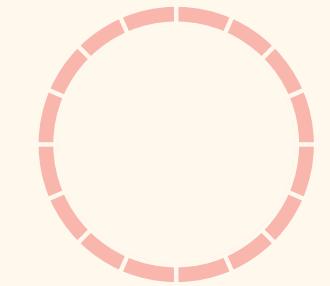
Logistic Regression

Decision Tree

Random Forest

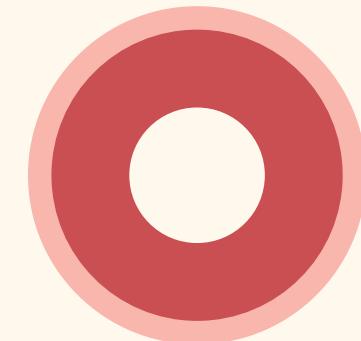
KNN

XGBoost



## Techniques

Baseline Model



## Hyperparameters

Optimized Models  
using Hyperparameters



TIP

Remember, **HadHeartAttack** is the Target

# BASELINE TEST PERFORMANCE EVALUATION

## Baseline Logistic Regression

**YES = 24%**

**YES = F1 SCORE - 34%**



**TIP**

Remember, **Recall** is the Key Metric

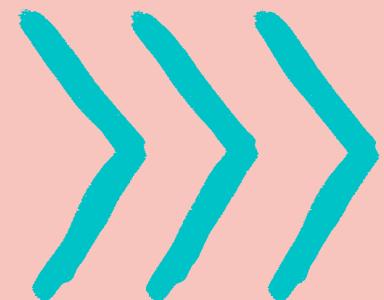
# TEST PERFORMANCE EVALUATION

Best Logistic  
Regression

YES = 71%

YES = F1 SCORE - 37%

Threshold



0.065



TIP

Remember, **Recall** is the Key Metric

# TEST PERFORMANCE EVALUATION

## Decision Tree

**YES = 26%**

**YES = F1 SCORE - 34%**



**TIP**

Remember, **Recall** is the Key Metric

# TEST PERFORMANCE EVALUATION

## Random Forest

**YES = NA%**

**YES = F1 SCORE - NA**



**TIP**

Remember, **Recall** is the Key Metric

# TEST PERFORMANCE EVALUATION

KNN

**YES = 12%**

**YES = F1 SCORE - 18%**



**TIP**

Remember, **Recall** is the Key Metric

# TEST PERFORMANCE EVALUATION

XGBoost

YES = 24%

YES = F1 SCORE - 33%



TIP

Remember, **Recall** is the Key Metric

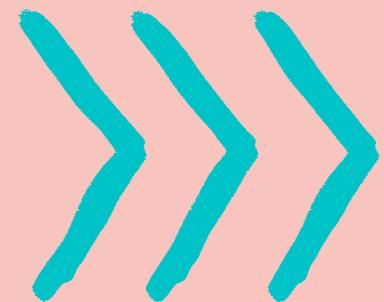
# TEST PERFORMANCE EVALUATION

Best XGBoost

YES = 73%

YES = F1 SCORE - 34%

Threshold



0.065



TIP

Remember, **Recall** is the Key Metric

# TEST PERFORMANCE EVALUATION

## Baseline Logistic Regression

**YES = 24%**

**YES = F1 SCORE - %**



**TIP**

## Best Logistic Regression

**YES = 71%**

**YES = F1 SCORE - 37%**

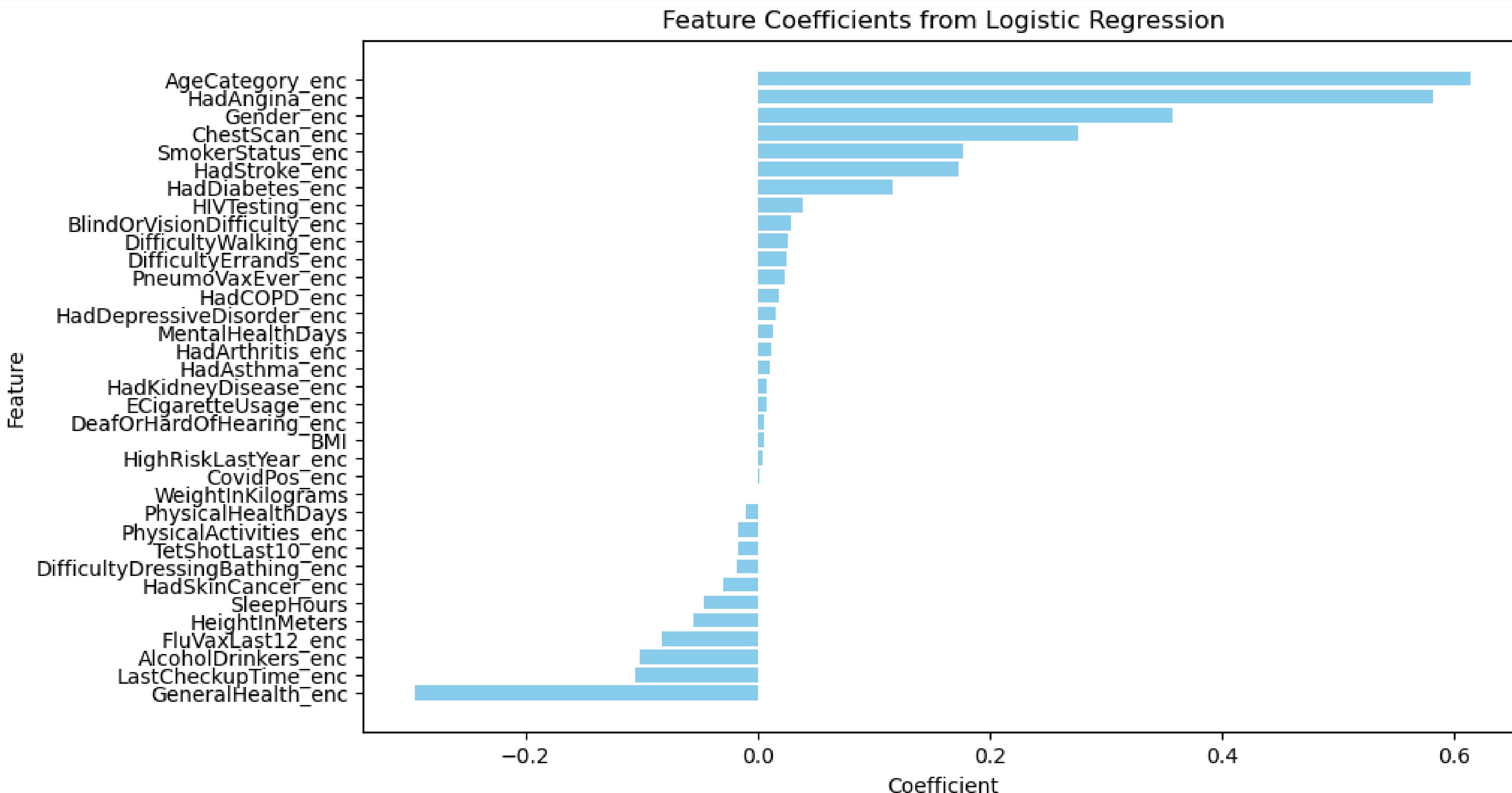
## Best XGBoost

**YES = 73%**

**YES = F1 SCORE - 34%**

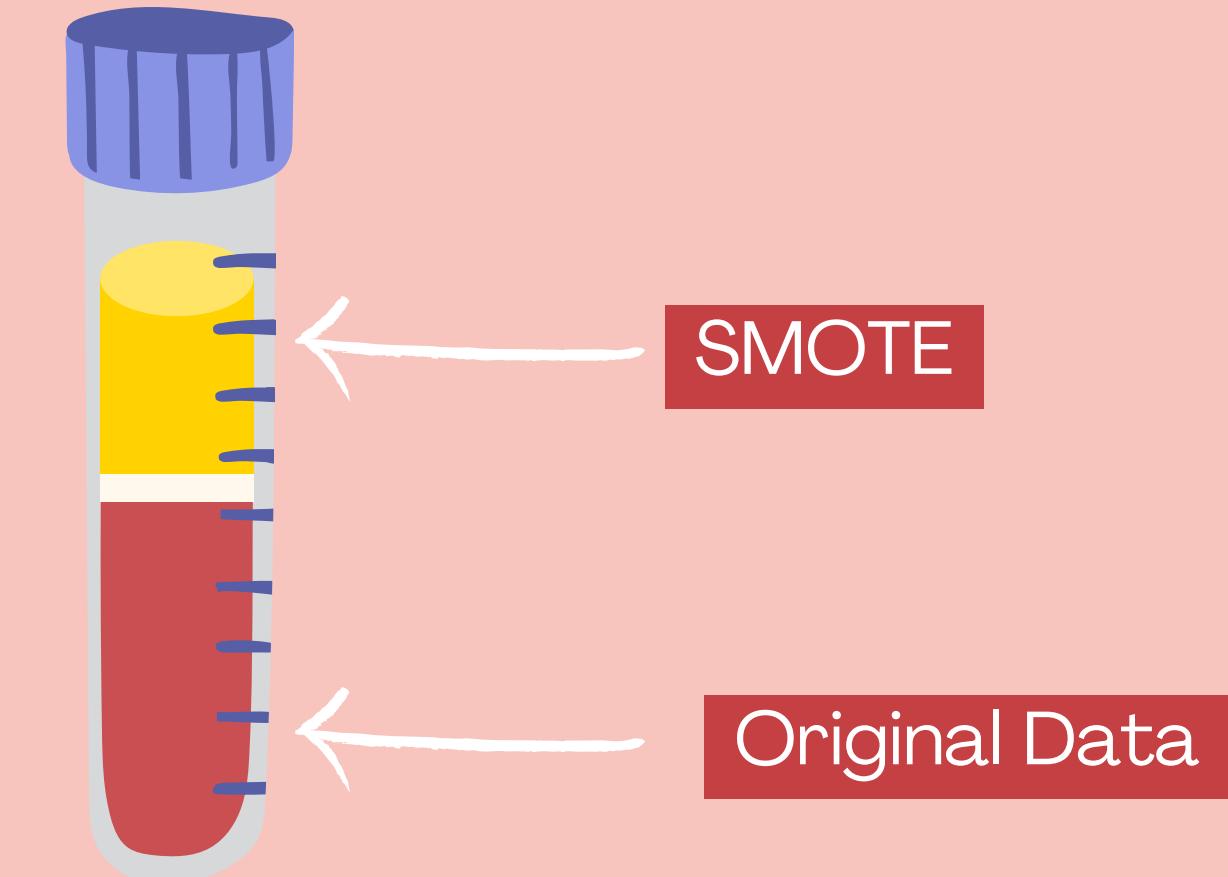
Remember, **Recall** is the Key Metric

# BEST LR FEATURE COEFFICIENTS



HadHeartAttack

# SMOTE



Smote adds synthetic data to an imbalanced feature.

TIP

Remember, **HadHeartAttack** is the Target

# SMOTE TEST PERFORMANCE EVALUATION

Smote Logistic  
Regression

**YES = 73%**

**YES = F1 SCORE - 35%**



**TIP**

Remember, **Recall** is the Key Metric



# HAD ANGINA

Data  
HadAngina

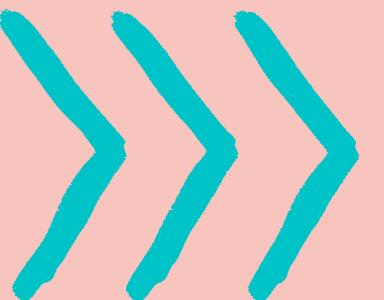
# HAR LR TEST PERFORMANCE EVALUATION

HAR Logistic  
Regression

**YES = 71%**

**YES = F1 SCORE - 37%**

Threshold



0.065



**TIP**

Remember, **Recall** is the Key Metric

# HAR FEATURE COEFFICIENTS

	Feature Coeffs	Coeffs
1	AgeCategoryenc	0.6
2	Gender_enc	0.36
3	ChestScan_enc	0.28
4	Smoker Status_enc	0.18
5	HadStroke_enc	0.17



**IMPACT**

so...



**SmoteNC**  
**Different Models**  
**More Evaluation**



**QUESTIONS?**

# 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