



MACHINE LEARNING

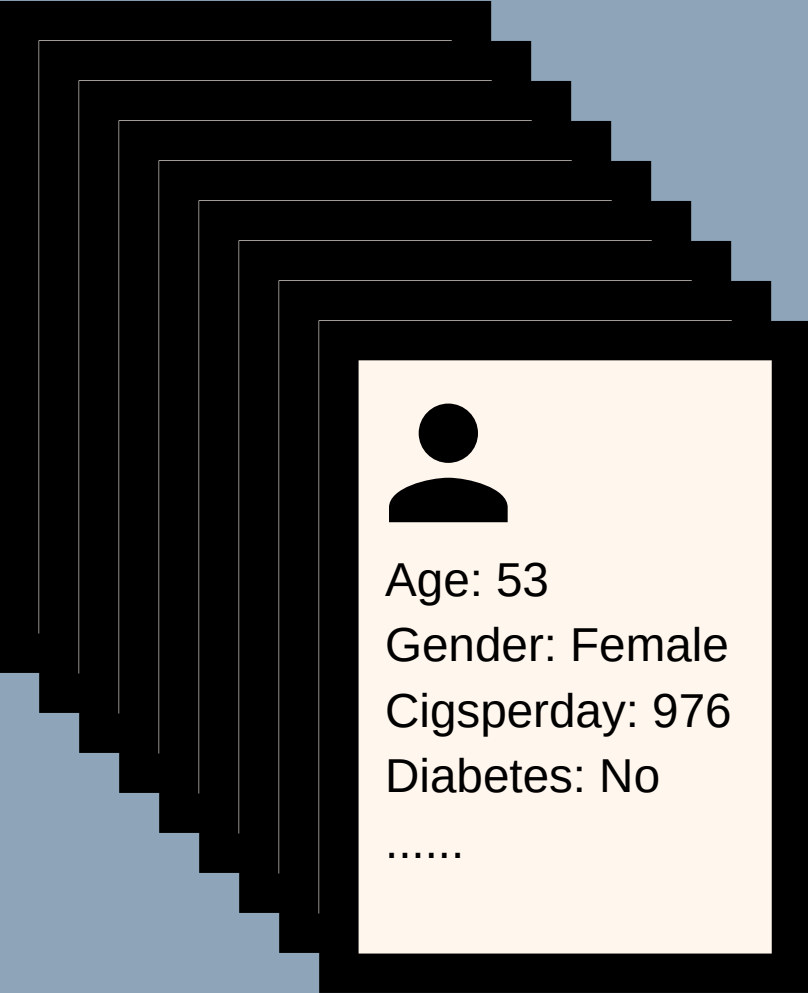
# Predicting Heart Disease

LAURA WÜRZ - OCTOBER 2019

# GOAL

PREDICT WETHER A  
PATIENT WILL DEVELOP  
A HEART DISEASE

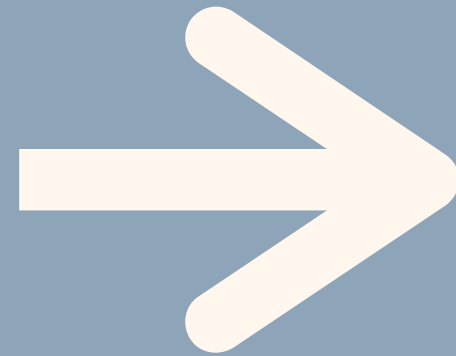
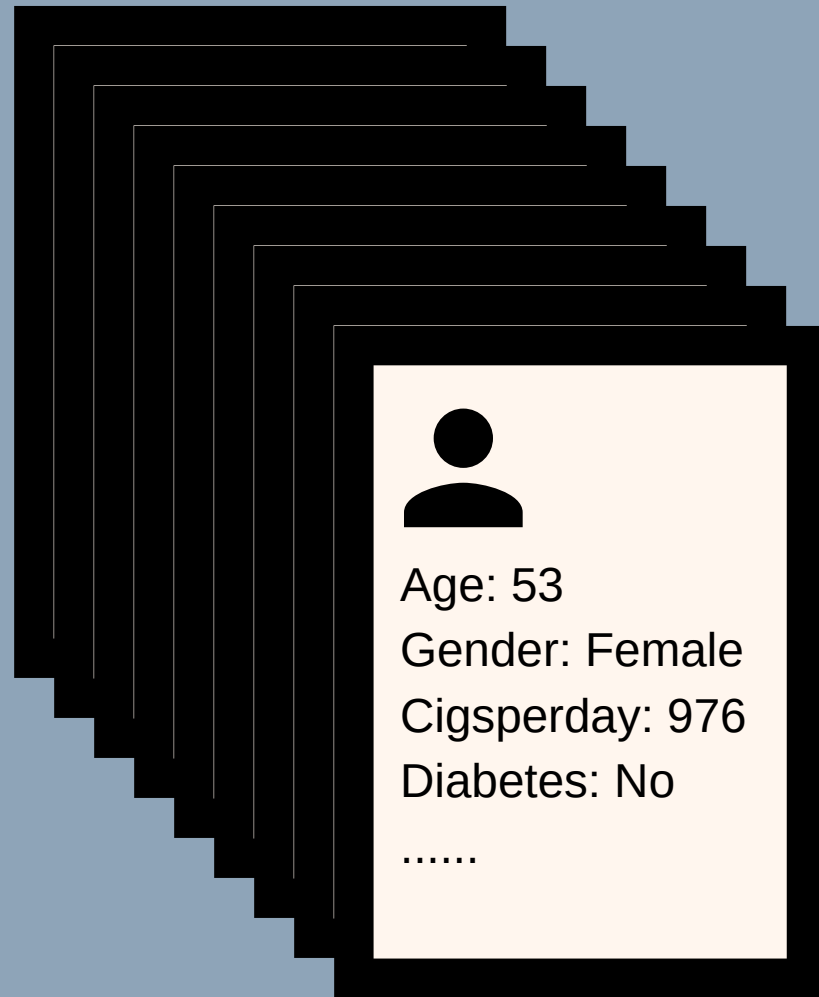
# DATA



Age: 53  
Gender: Female  
Cigsperday: 976  
Diabetes: No

.....

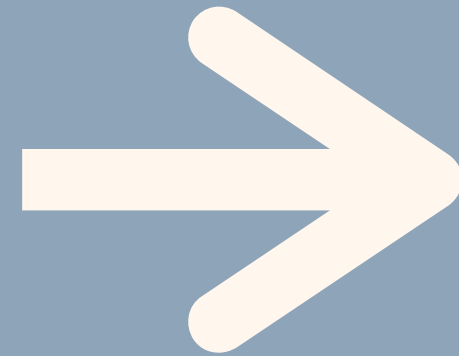
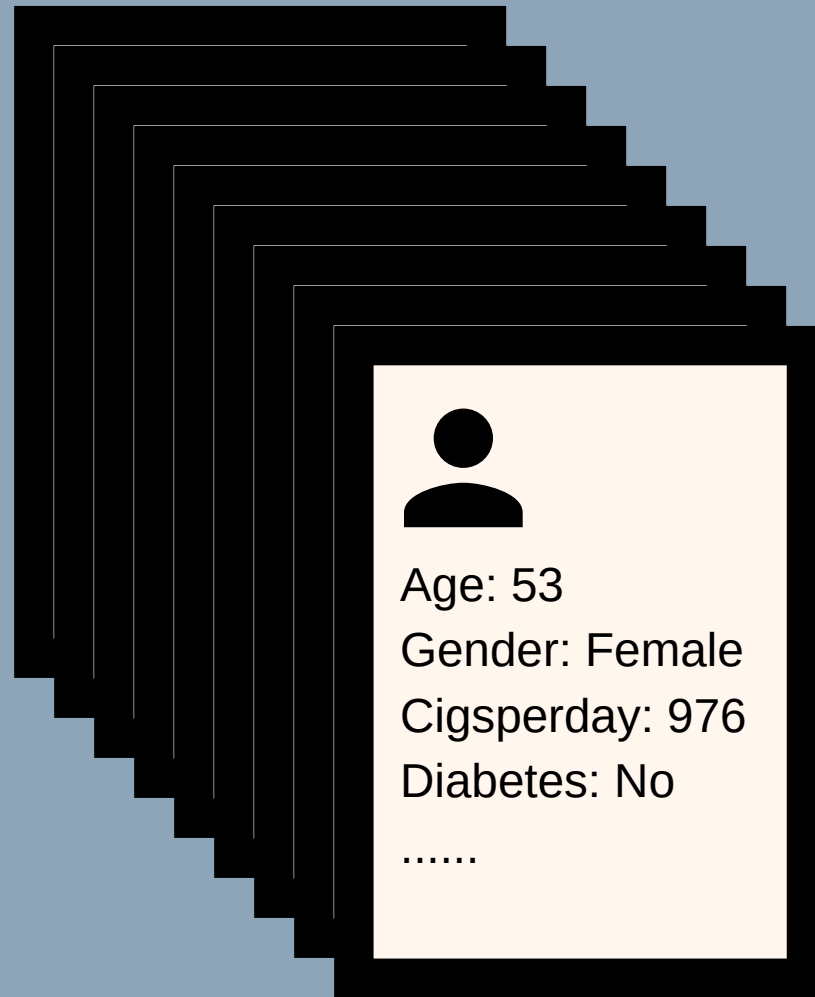
# DATA



# MODEL



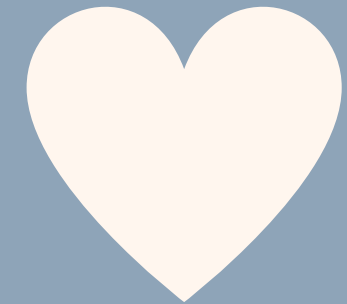
# DATA



# MODEL

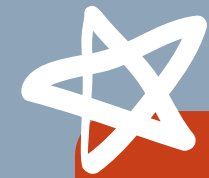


# PREDICTION



# WORKFLOW

---



EXPLORE & CLEAN  
DATASET

FEATURE  
SCALING

TEST - TRAIN  
SPLIT

EVALUATION

FIT MODELS

RESAMPLING

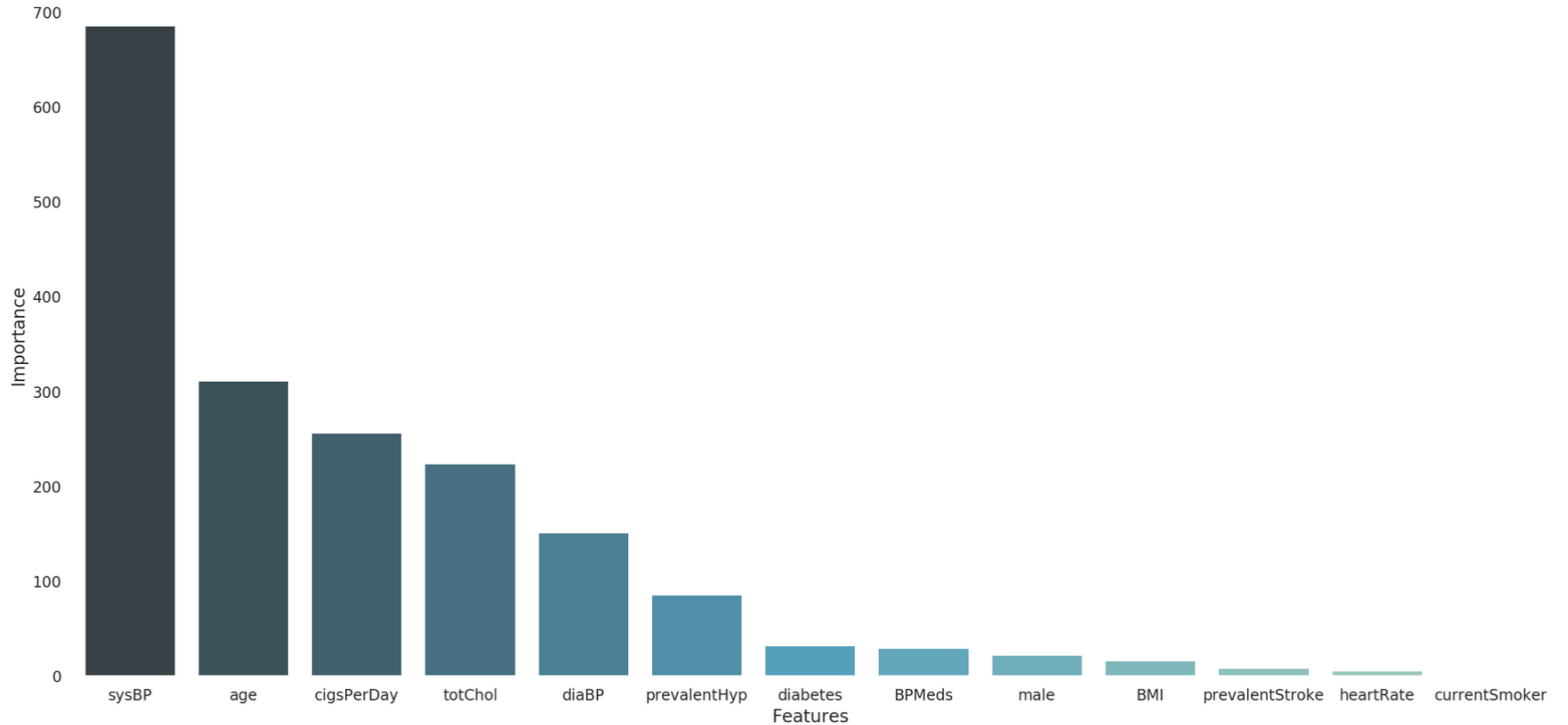
CHOOSE MODEL



EXPLORE & CLEAN  
DATASET

# FEATURE SELECTION

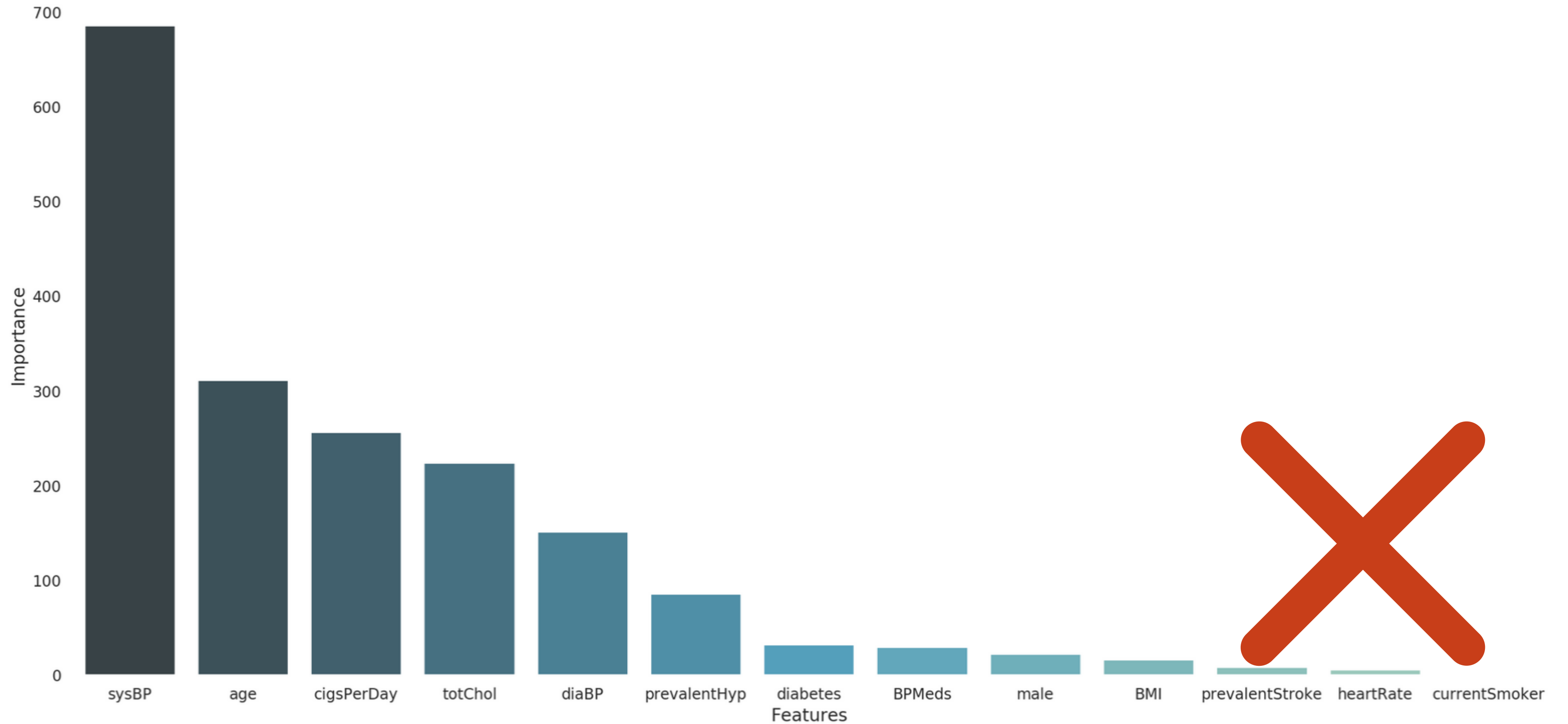
Feature importance



EXPLORE & CLEAN  
DATASET

# FEATURE SELECTION

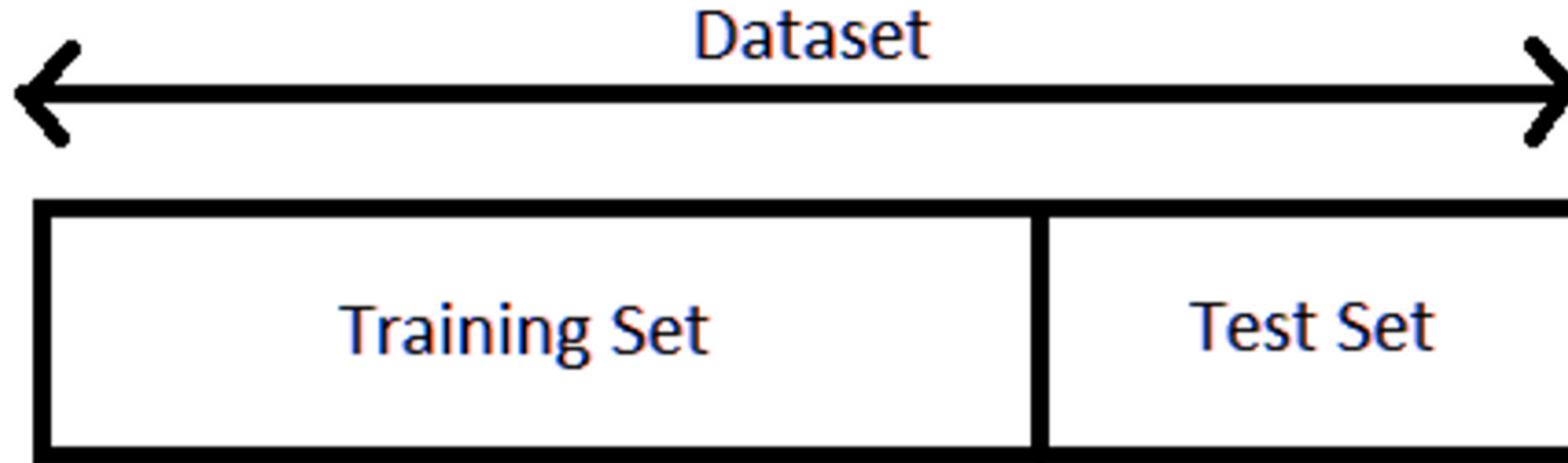
Feature importance





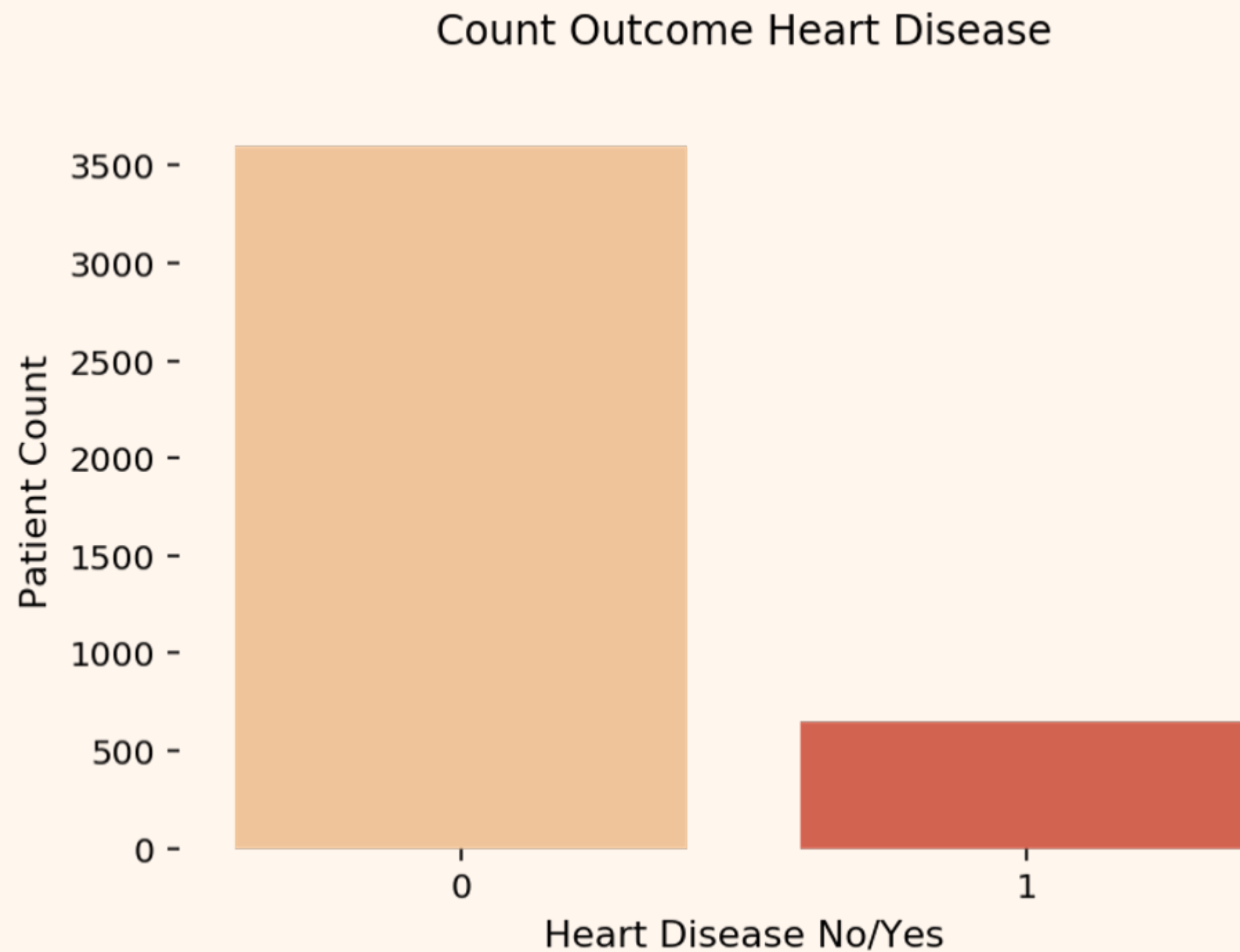
TEST - TRAIN  
SPLIT

# TEST-TRAIN SPLIT

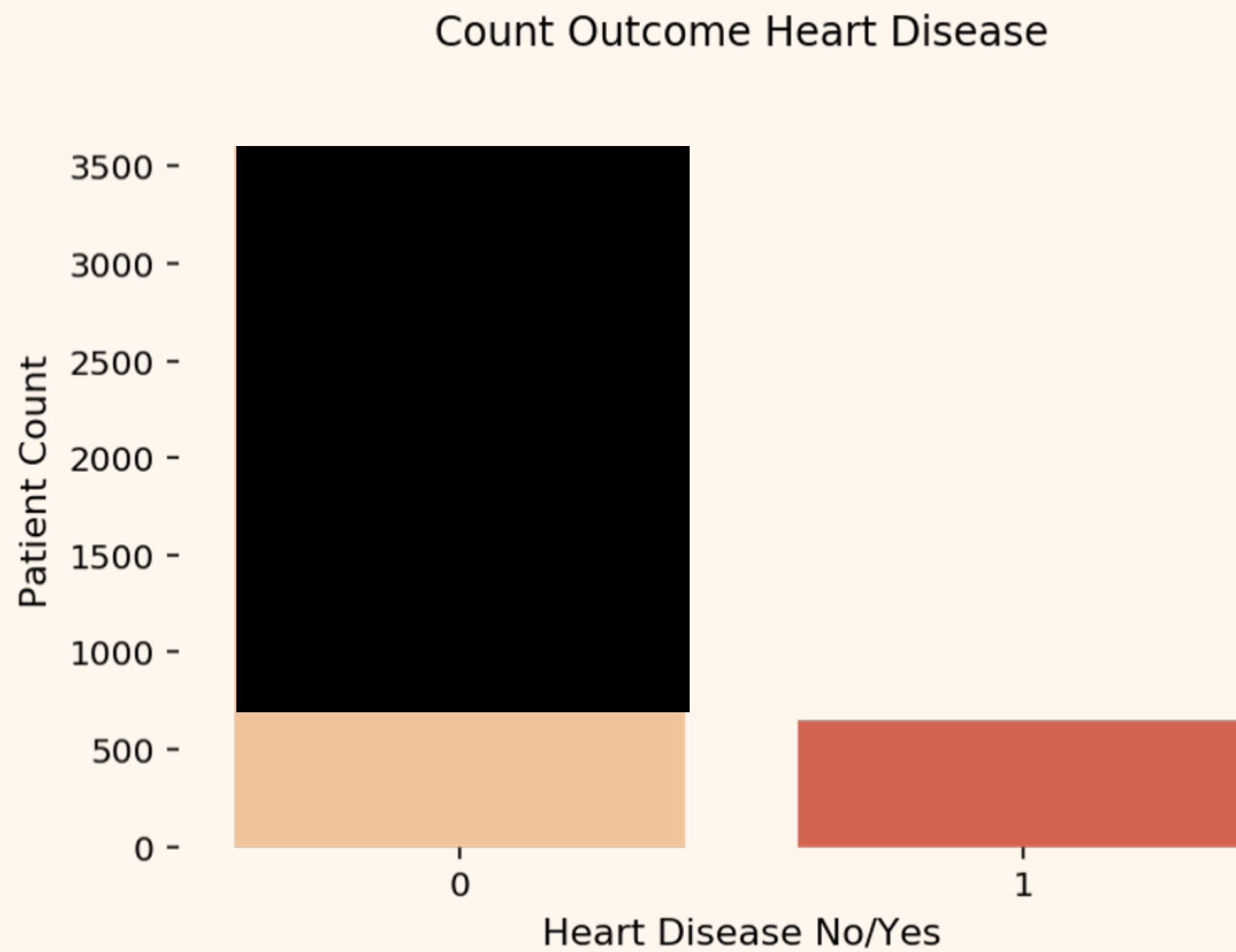


RESAMPLING

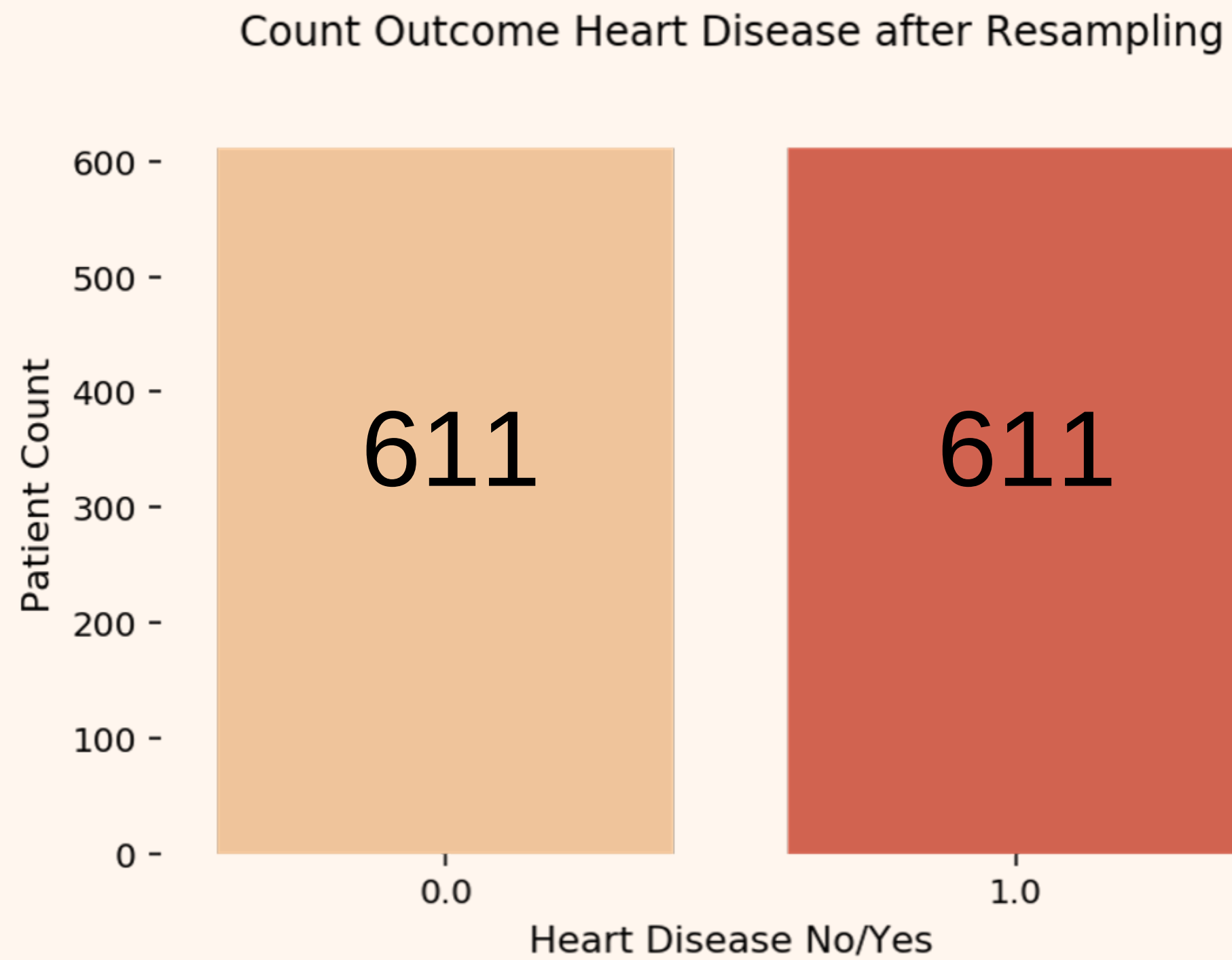
# UNBALANCED DATASET PROBLEM



# UNDERSAMPLING



# RESAMPLING



FIT MODELS

# CLASSIFICATION ALGORITHMS



LOGISTIC  
REGRESSION

SUPPORT VECTOR  
MACHINE

DECISION TREE

K-NEAREST  
NEIGHBOUR

FIT MODELS

# MODEL ALGORITHMS



LOGISTIC  
REGRESSION

66.3%

SUPPORT  
VECTOR  
MACHINE

64.3 %

DECISION  
TREE

70.4 %

K-NEAREST  
NEIGHBOUR

80.9 %

FIT MODELS

# MODEL ALGORITHMS



LOGISTIC  
REGRESSION

66.3%

SUPPORT  
VECTOR  
MACHINE

64.3 %

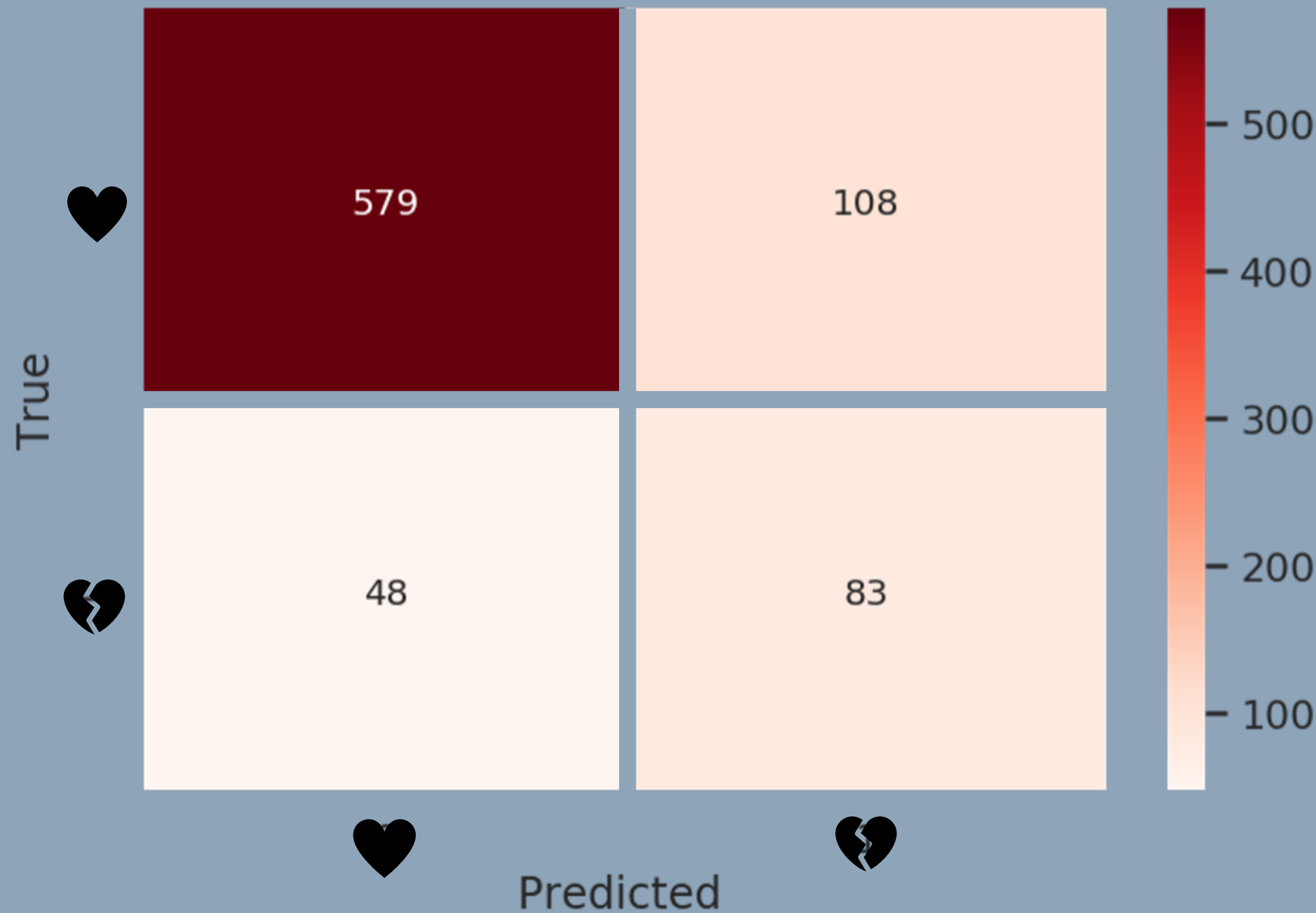
DECISION  
TREE

70.4 %

K-NEAREST  
NEIGHBOUR

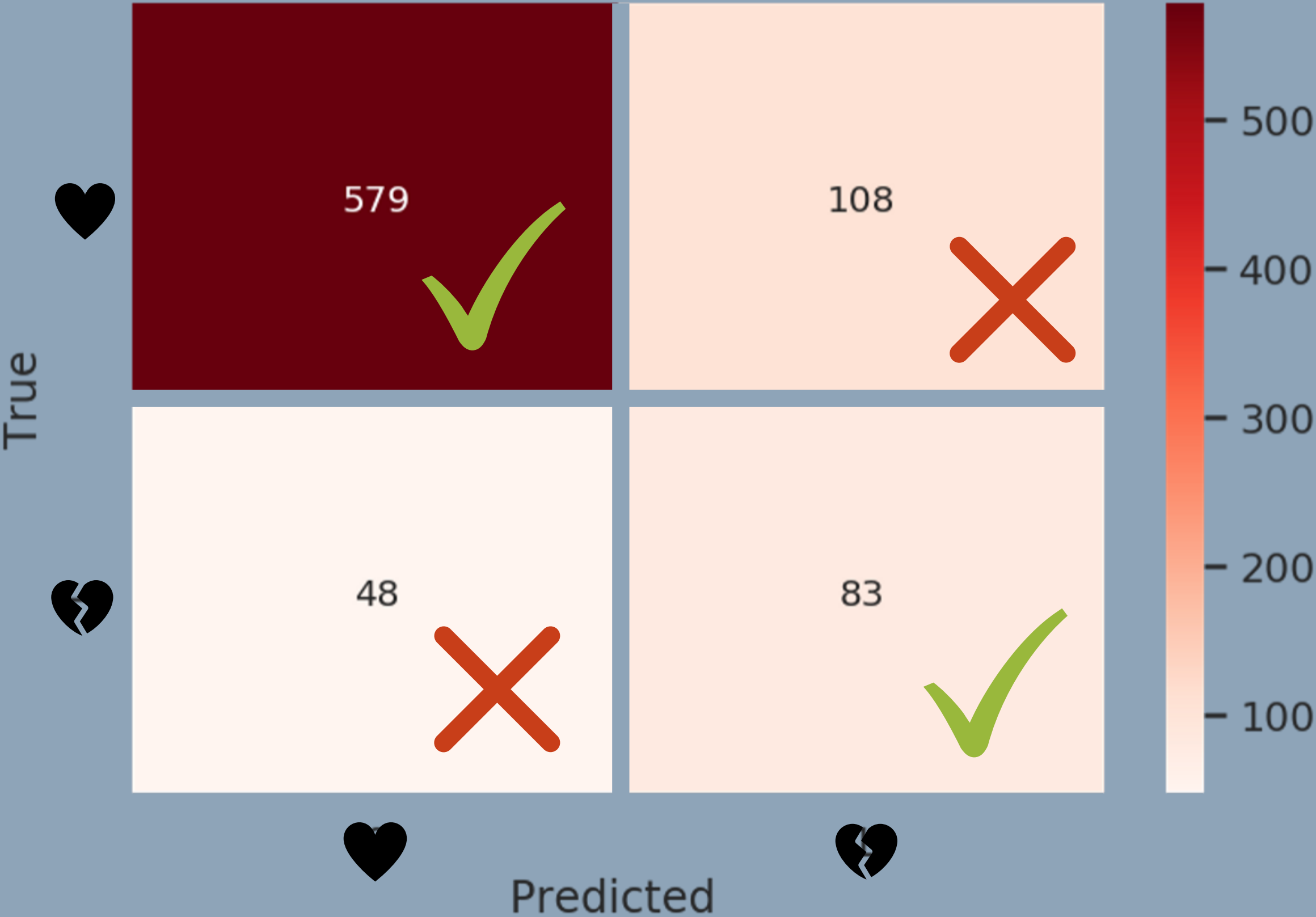
80.9 %

# CONFUSION MATRIX KNN

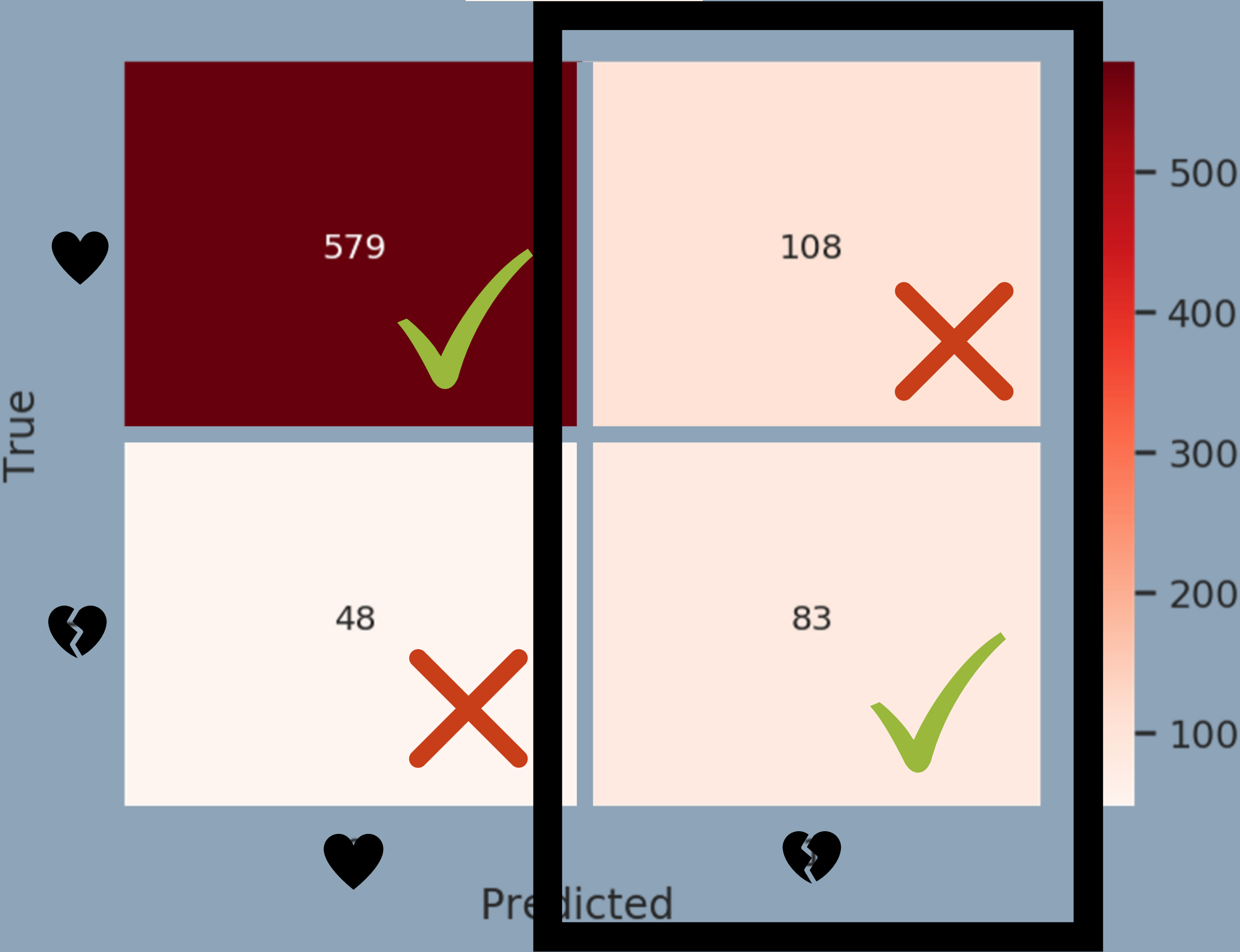




# CONFUSION MATRIX KNN

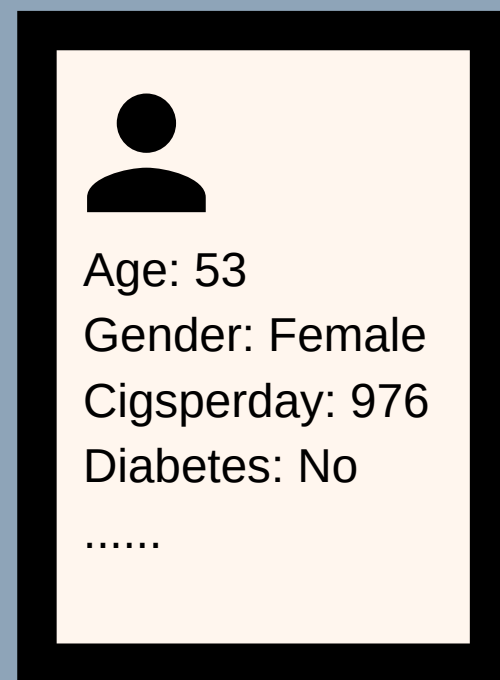


# CONFUSION MATRIX KNN

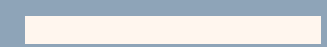


# APPLY MODEL IN THE REAL WORLD

---



THANK YOU!



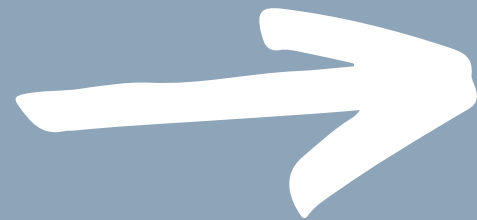
Questions?



## FEATURE SCALING

# FEATURE SCALING

	sysBP	age	cigsPerDay
count	4088.000000	4088.000000	4088.000000
mean	132.213796	49.498532	8.997554
std	21.899294	8.544161	11.915472
min	83.500000	32.000000	0.000000
25%	117.000000	42.000000	0.000000
50%	128.000000	49.000000	0.000000
75%	143.500000	56.000000	20.000000
max	295.000000	70.000000	70.000000



	sysBP	age	cigsPerDay
count	4088.000000	4088.000000	4088.000000
mean	0.230325	0.460488	0.128536
std	0.103543	0.224846	0.170221
min	0.000000	0.000000	0.000000
25%	0.158392	0.263158	0.000000
50%	0.210402	0.447368	0.000000
75%	0.283688	0.631579	0.285714
max	1.000000	1.000000	1.000000

## AUC ROC CURVE KNN

ROC CURVE KNN

