

Photo by Alexandru Acea on Unsplash

CLASSIFICATION PROJECT

PREDICTING HEART DISEASE

Sandra Paredes

INTRODUCTION

- ▶ **Motivation:** Kaiser Permanente, an HMO, wants to identify patients at high risk for heart disease who would benefit from a heart health program.
- ▶ **Research Question:** How might we predict which patients are at high risk of heart disease?
- ▶ **Impact Hypothesis:** Reduce the number of patients who develop heart disease (arterial plaque or heart attack).

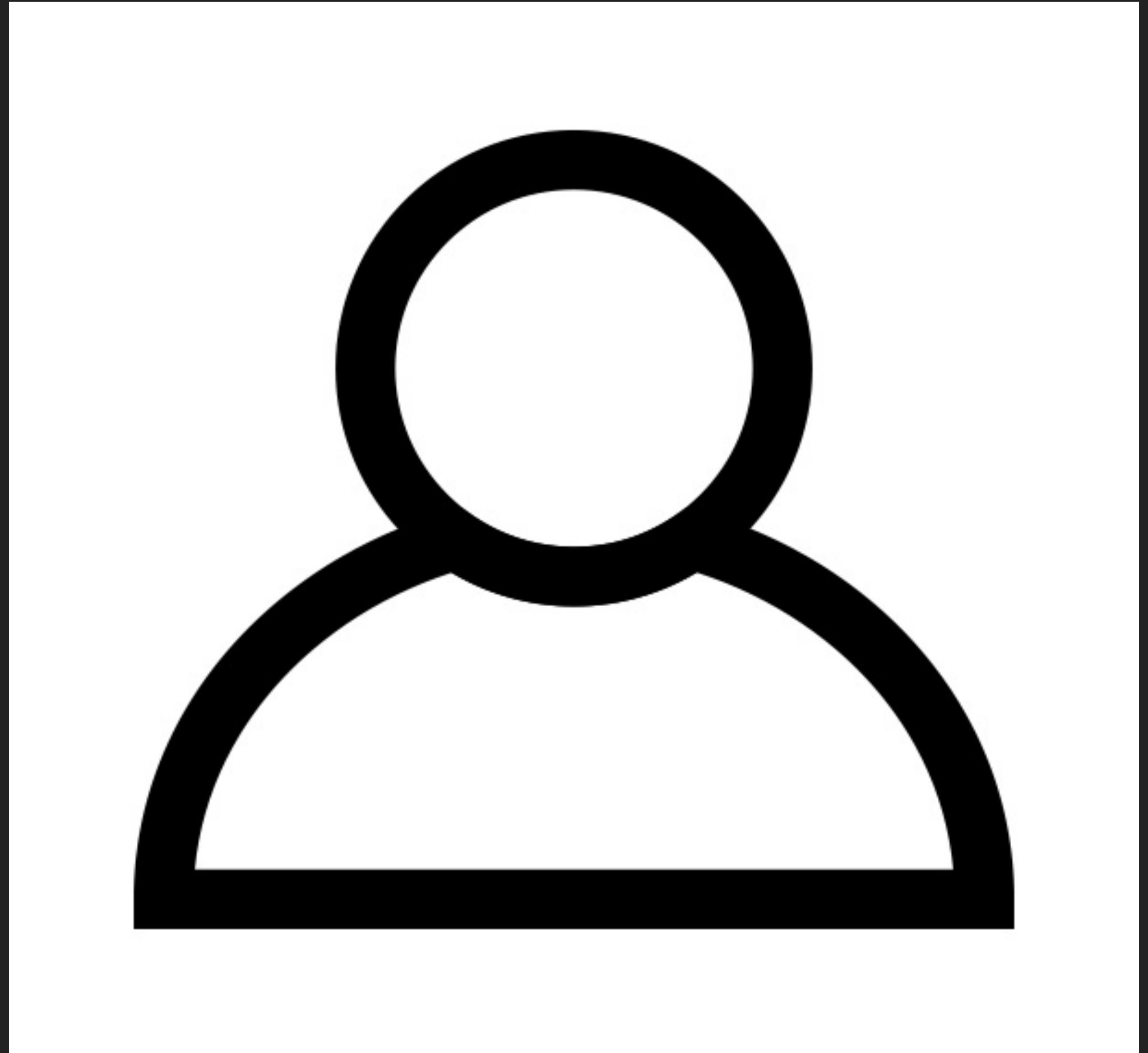


Photo by

METHODOLOGY

- ▶ Dataset: Kaggle Personal Key Indicators of Heart Disease [2]
 - ▶ Excerpted from CDC BRFSS, 2020 [3]
 - ▶ Survey of Americans (n= 319,795)
- ▶ Target: **Heart Disease**
 - ▶ coronary heart disease
 - ▶ myocardial infarction

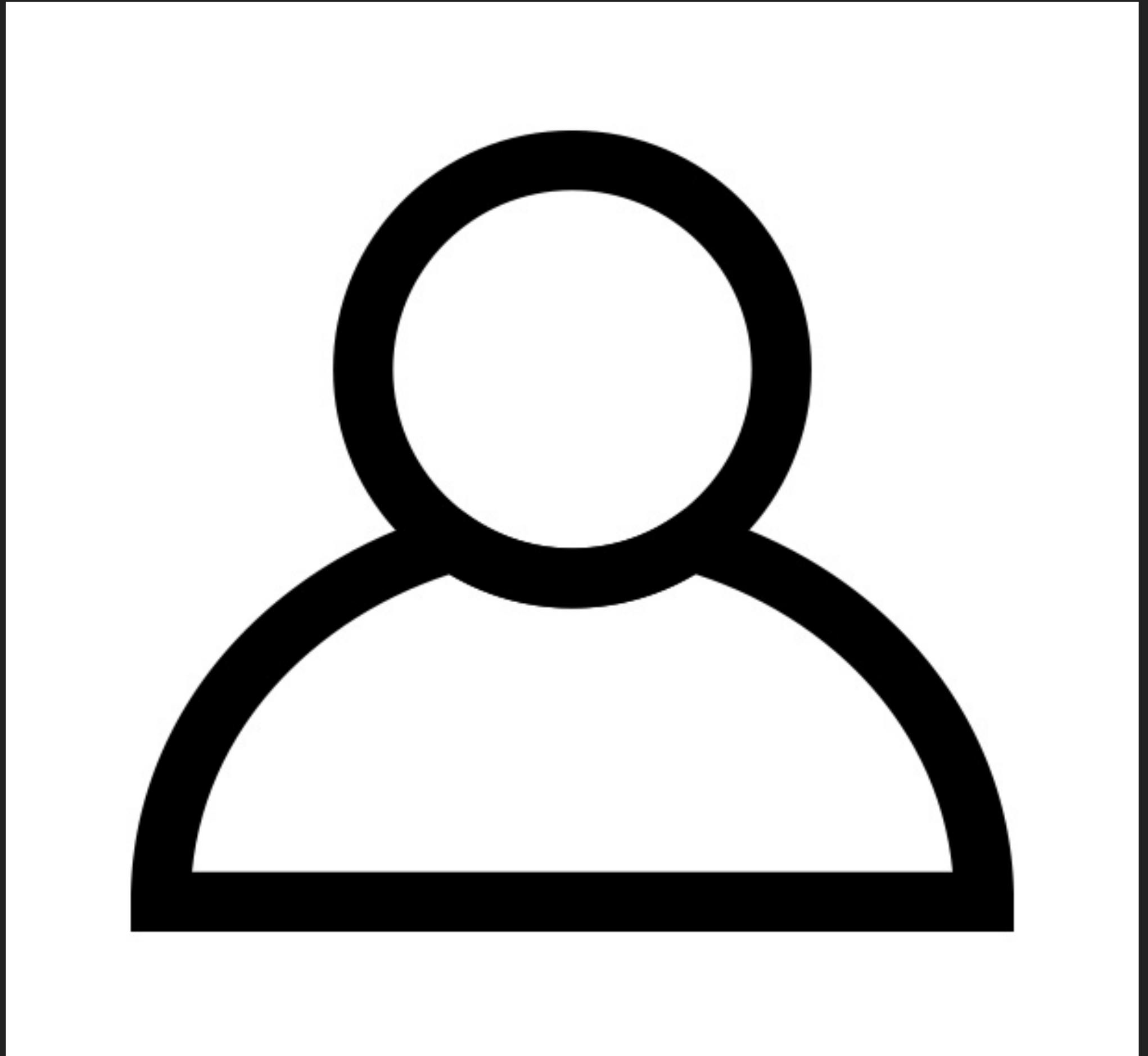


Photo by

METHODOLOGY

Transformations:

- ▶ Mapped categorial data
- ▶ Dummy variables

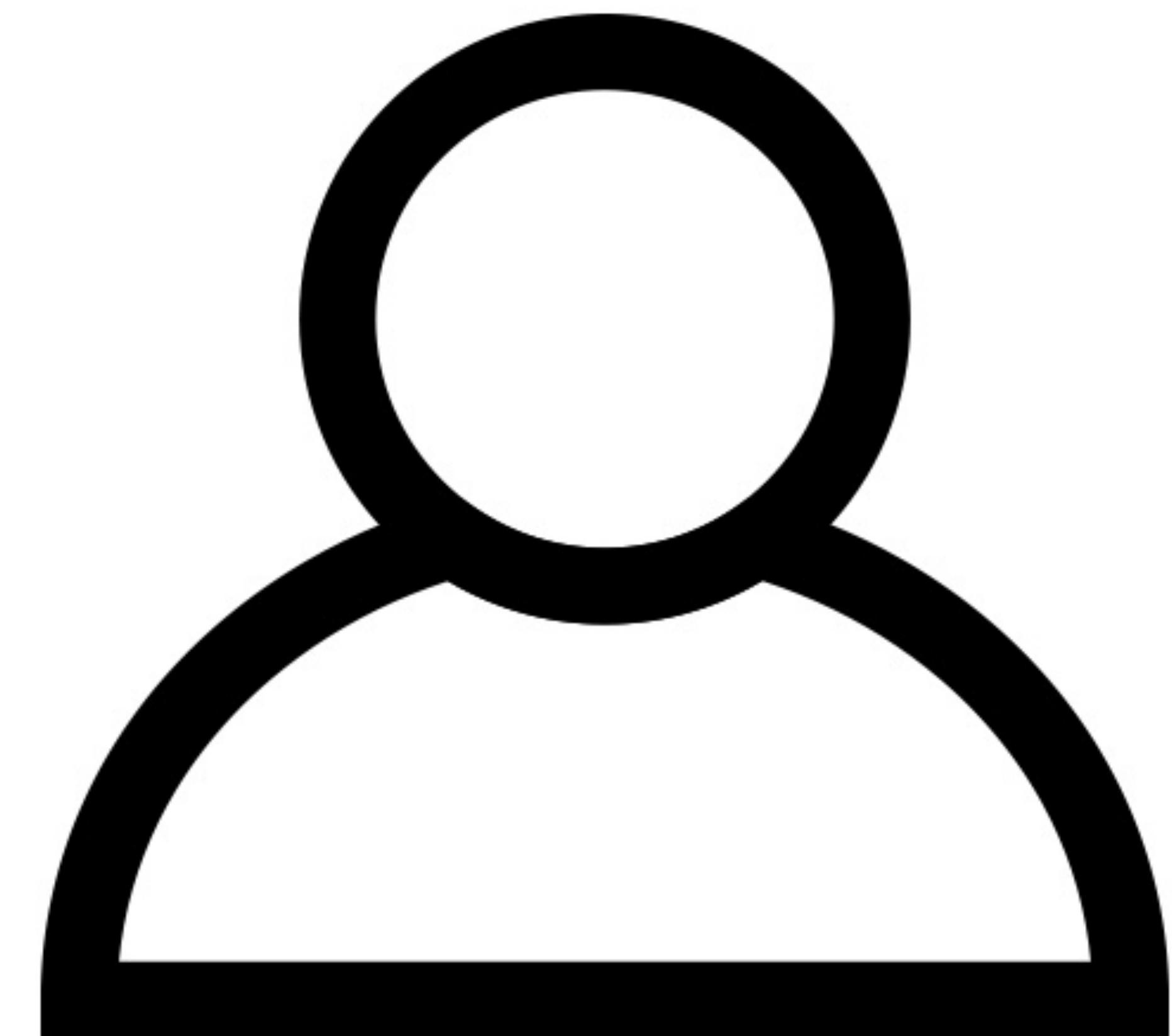
Feature engineering:

- ▶ Grouped demographics, behaviors, measures, and disease

Class imbalance handling:

- ▶ Up/down ... weight

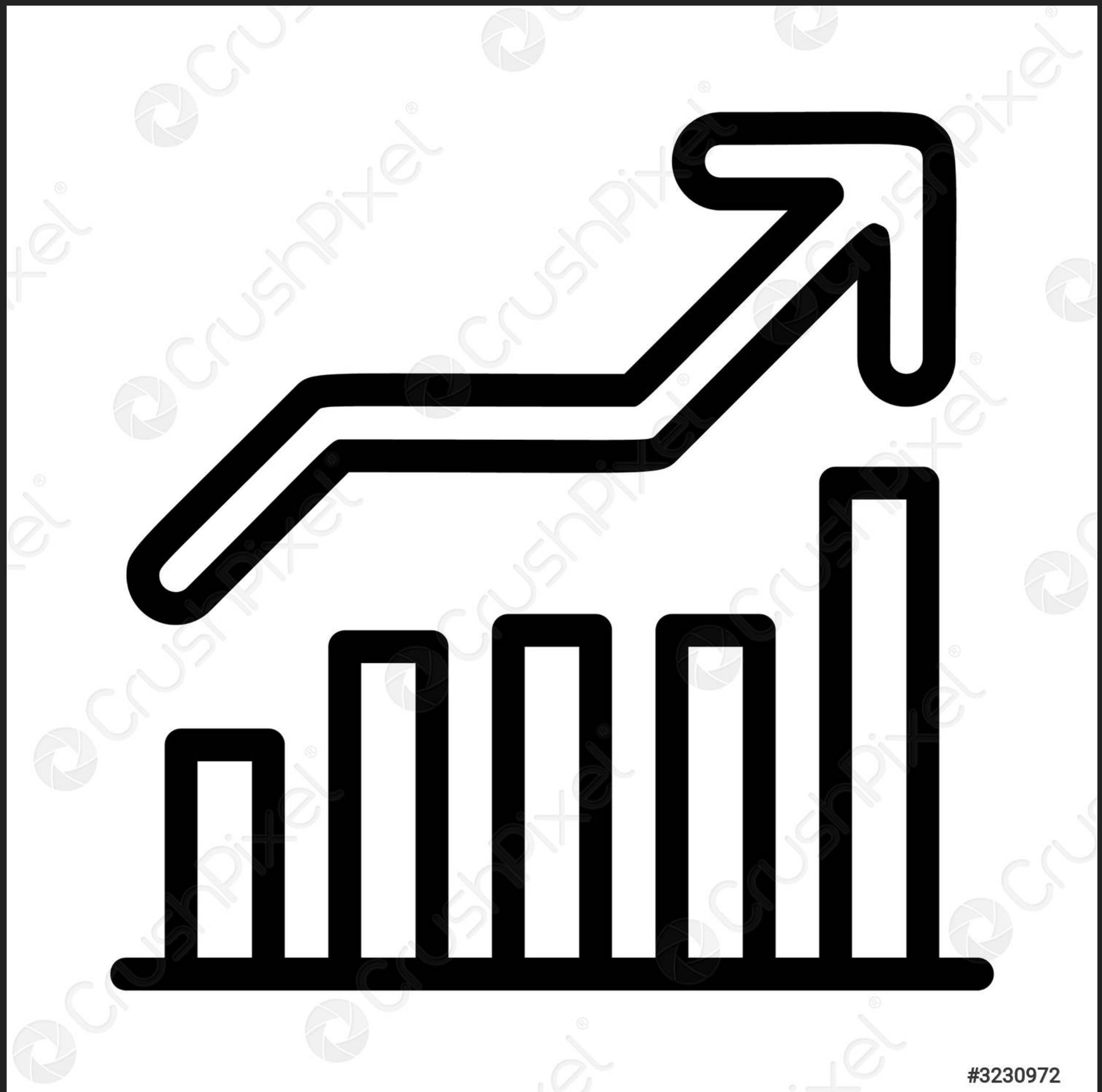
Figure ##



RESULTS

- ▶ Best model
- ▶ Error metric
- ▶ Score on (specify on what subset)
- ▶ Confusion matrix final model

Figure ##



RESULTS

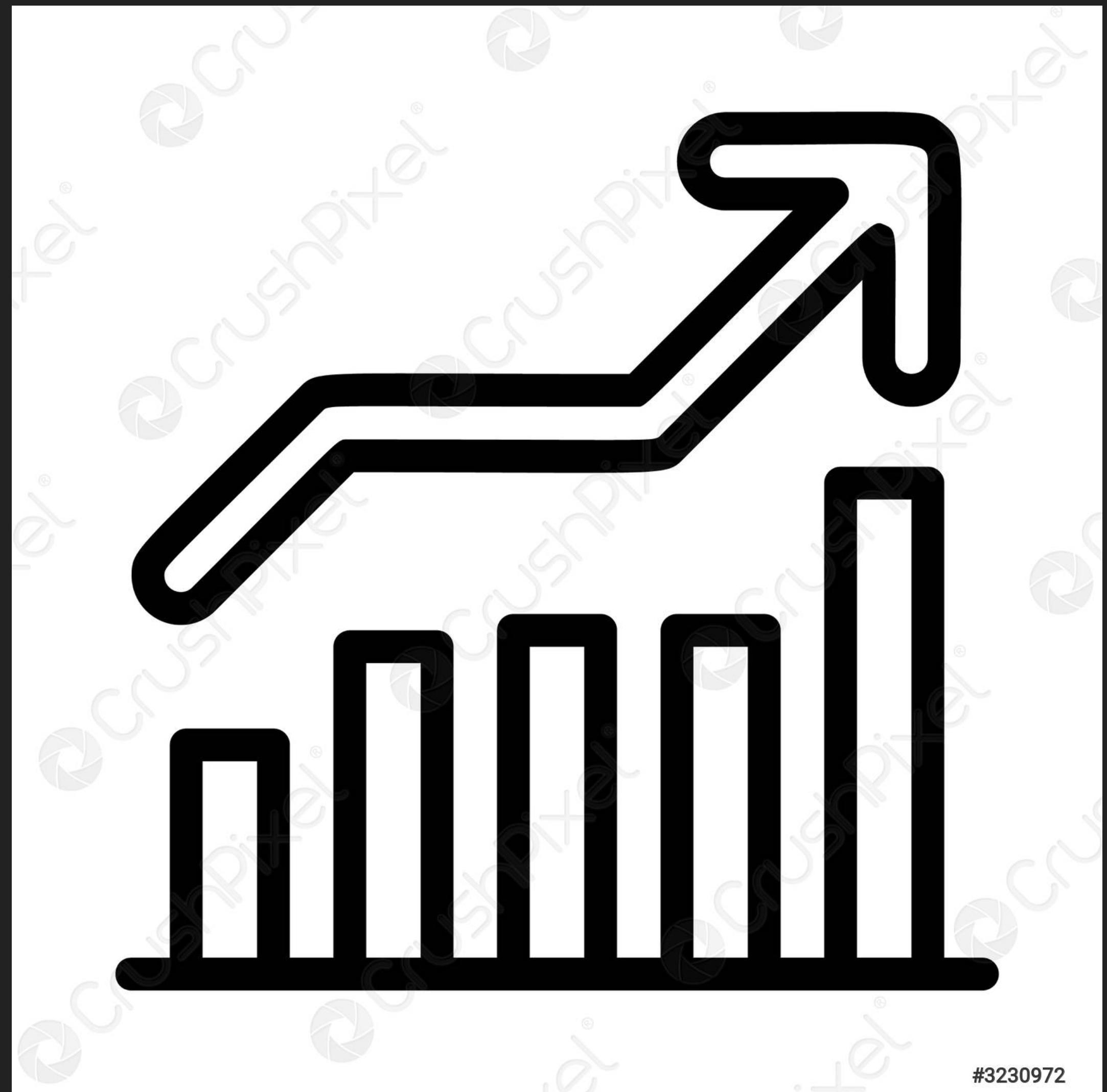
Recall (X_{test_scaled} , y_{test})	Score
Baseline (age, mobility, physical health)	0.3634
Logistic Regression	0.5327
Random Forest	0.3621
Engineered Features (behaviors, demographics, disease, health measures)	0.4731
Original + Engineered Features	0.5327

RESULTS

Top 3 features of importance:

- ▶ ...
- ▶ ...
- ▶ ...

Figure ##



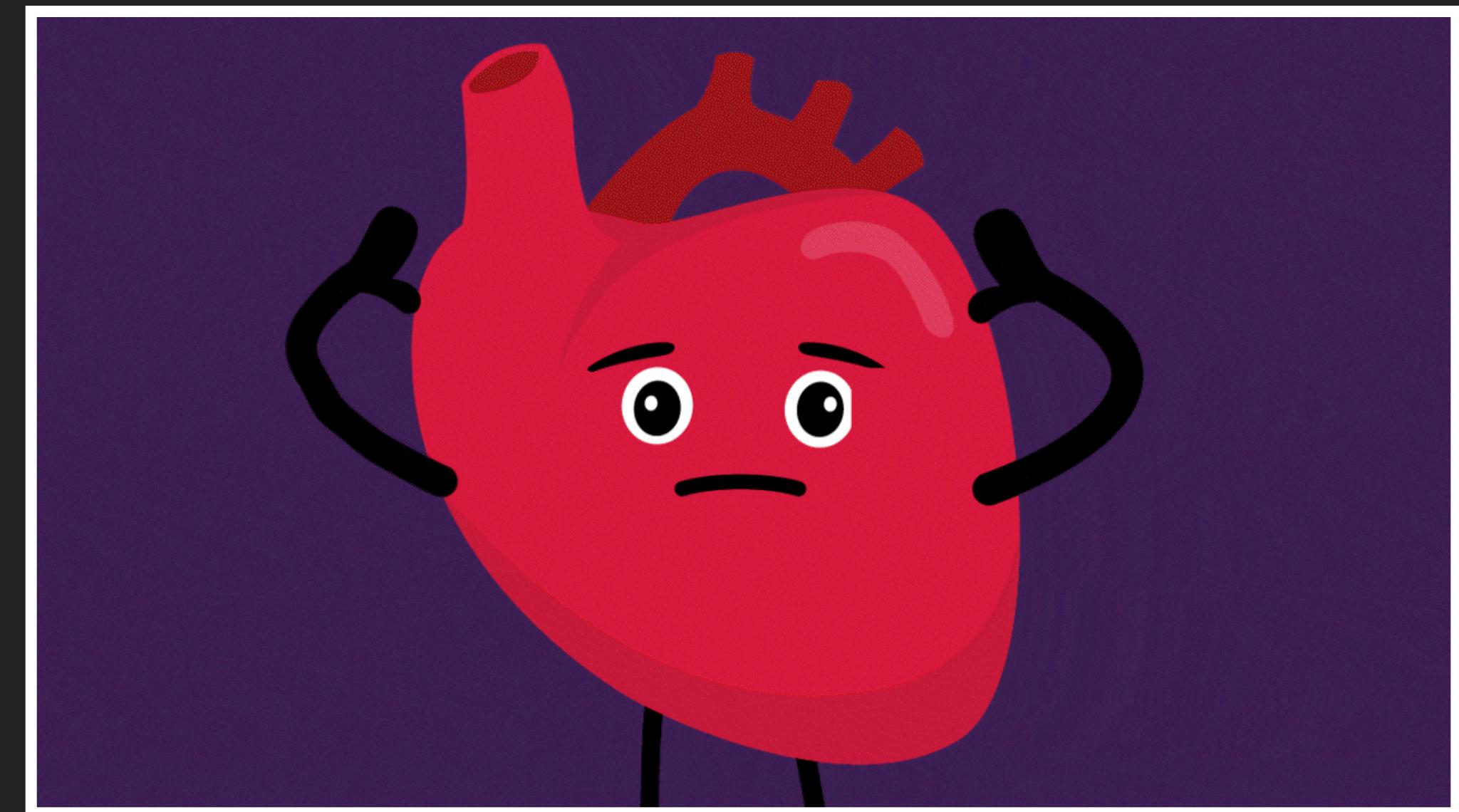
CONCLUSIONS

Insights

- ▶ ...
- ▶ ...
- ▶ ...

Recommendations

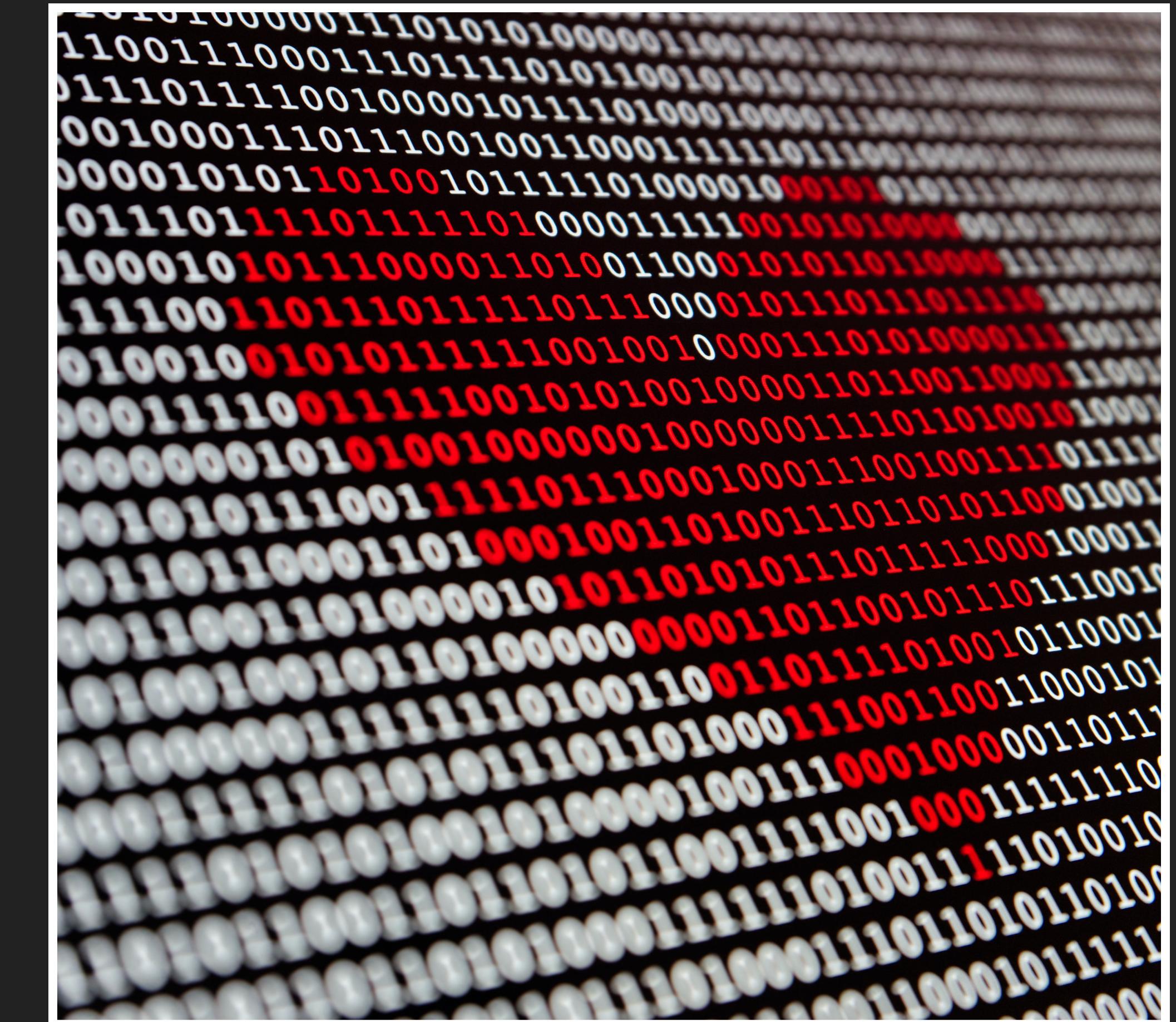
- ▶ ...
- ▶ ...
- ▶ ...



GIF by NHLBI #OurHearts

FUTURE WORK

- ▶ ...
- ▶ ...
- ▶ ...



[Photo by Alexander Sinn on Unsplash](#)

APPENDIX

- ▶ Summary, data, and slides are available at github.com/slp22/classification-project



[Photo by Kelly Sikkema on Unsplash](#)

BRFSS HEART DISEASE DATA

THANK YOU

APPENDIX: SOURCES

1. National Heart, Lung, and Blood Institute, Coronary Heart Disease: <https://www.nhlbi.nih.gov/health/coronary-heart-disease/causes>
2. Kaggle Personal Key Indicators of Heart Disease: <https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease>
3. CDC Behavioral Risk Factor Surveillance System: <https://www.cdc.gov/brfss/index.html>
4. Data Dictionary: https://www.cdc.gov/brfss/annual_data/2020/pdf/codebook20_llcp-v2-508.pdf

APPENDIX: DATA DICTIONARY

Target

- `y_heart_disease`: Y/N | coronary heart disease (CHD) or myocardial infarction (MI)

Health Behaviors

- `behavior_activity`: Num (0-30) | # days did physical activity/exercise other than regular job
- `behavior_alcohol`: Y/N | heavy drinker, defined as men: 14+/wk, women: 7+/wk (includes beer, wine, malt beverage, liquor)
- `behavior_sleep`: Num (0-24) | # hours of sleep in a 24-hour period, on average
- `behavior_tobacco`: Y/N | smoked at least 100 cigarettes in your life

Demographics

- `demg_age`: 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+
- `demg_gender`: male/female
- `demg_race`: White, Black, Asian, American Indian/Alaskan Native, Hispanic, Other race

Health Measures

- `health_bmi`: Num | Body Mass Index (BMI)
- `health_physical`: Num (0-30) | # days physical health not good, includes physical illness and injury
- `health_mental`: Num (0-30) | # days mental health not good, includes stress, depression, and problems with emotions
- `health_general`: Excellent, Very Good, Fair, Poor | Would you say that in general your health is...
- `health_mobility`: Y/N | serious difficulty walking or climbing stairs

Chronic Disease

- `disease_asthma`: Y/N
- `disease_diabetes`: Y/N/Y pregnancy/N borderline
- `disease_kidney`: Y/N | kidney disease, excludes kidney stones, bladder infection or incontinence
- `disease_skin`: Y/N | skin cancer
- `disease_stroke`: Y/N