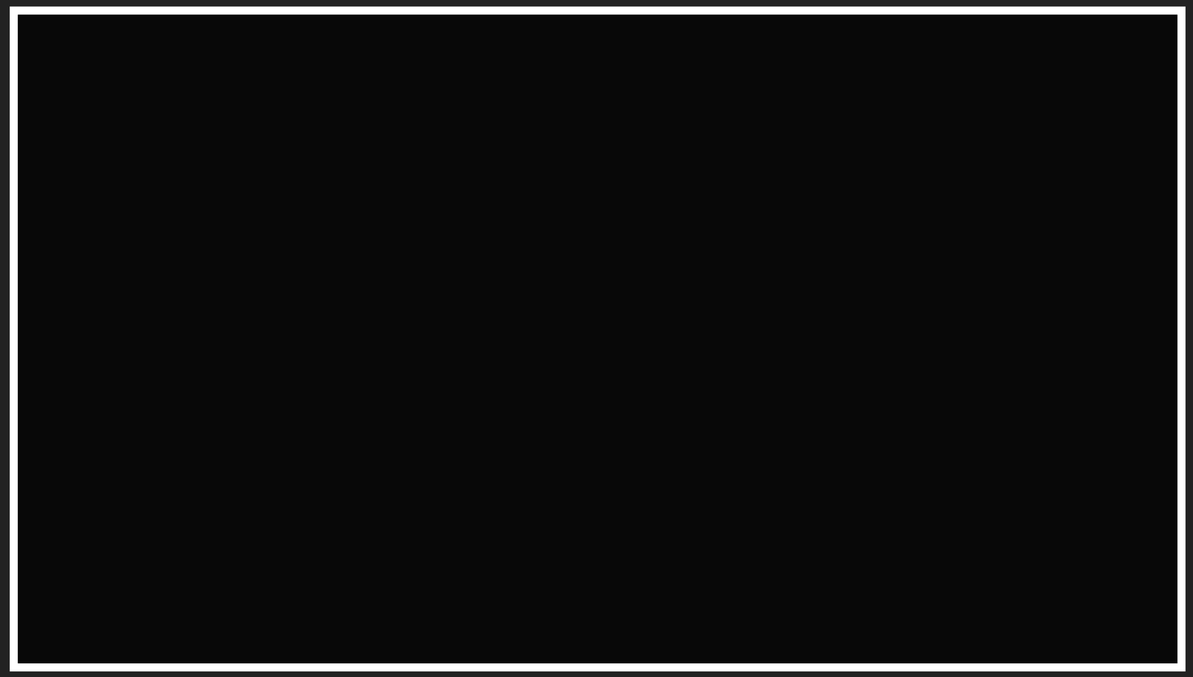


GIF by NHLBI #OurHearts

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).

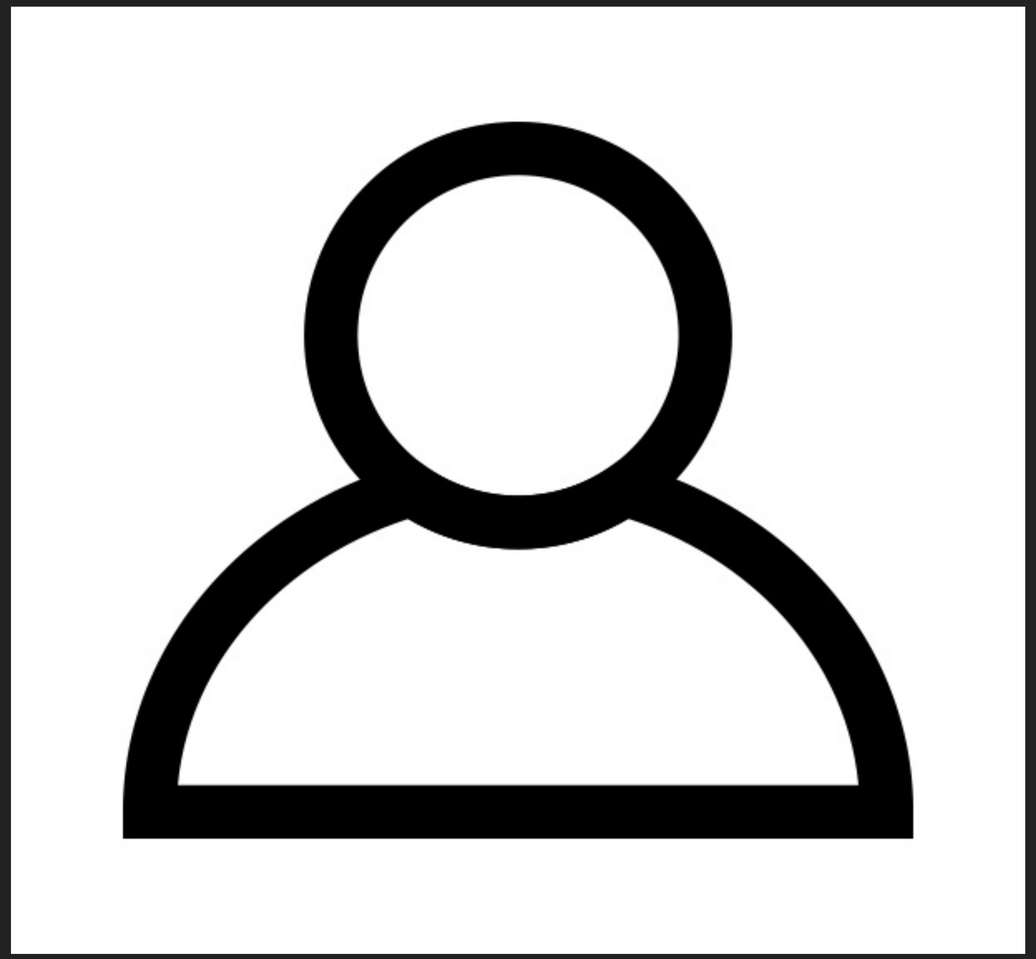


Photo by

METHODOLOGY

- Dataset
  - Kaggle, Indicators of Heart Disease [2]
  - Excerpted from CDC BRFSS, 2020 [3]
  - Survey of American adults
  - n= 319,795
- Target
  - Heart Disease
    - coronary heart disease
    - myocardial infarction

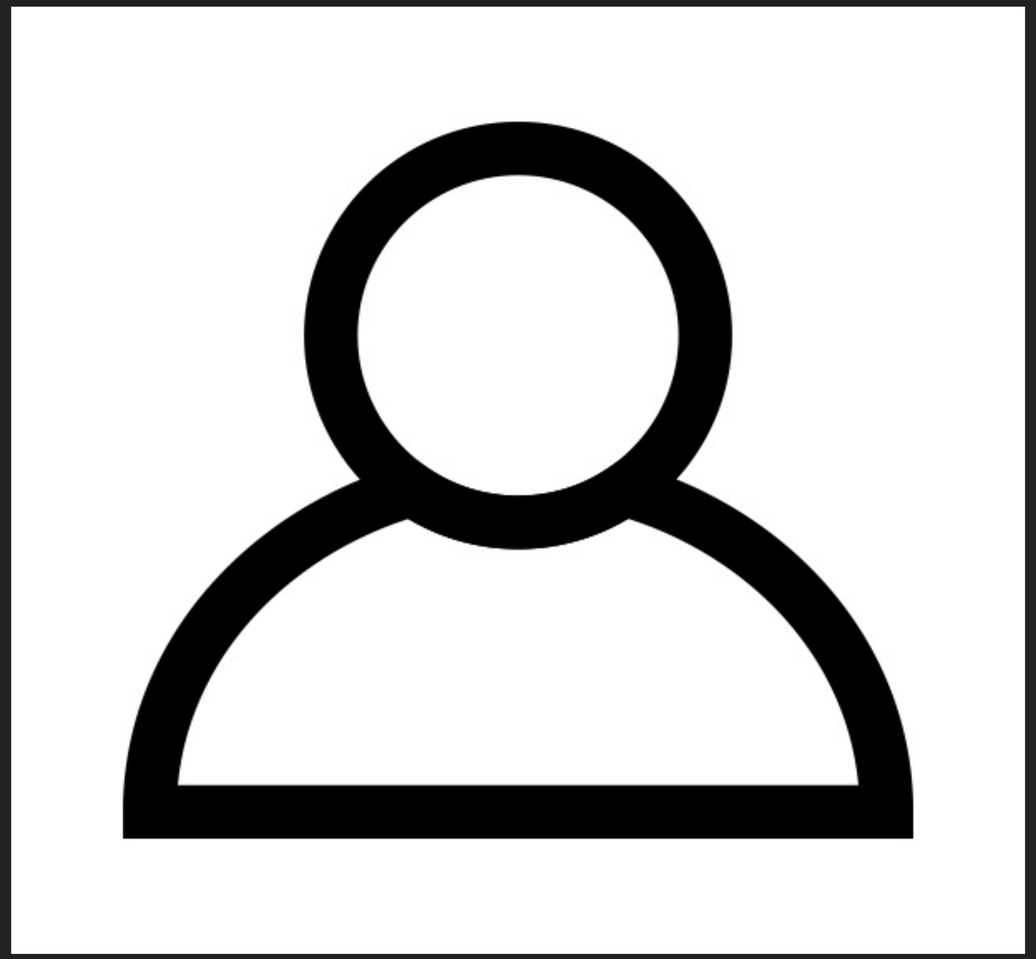


Photo by

METHODOLOGY

Transformation

- Mapped categorical values

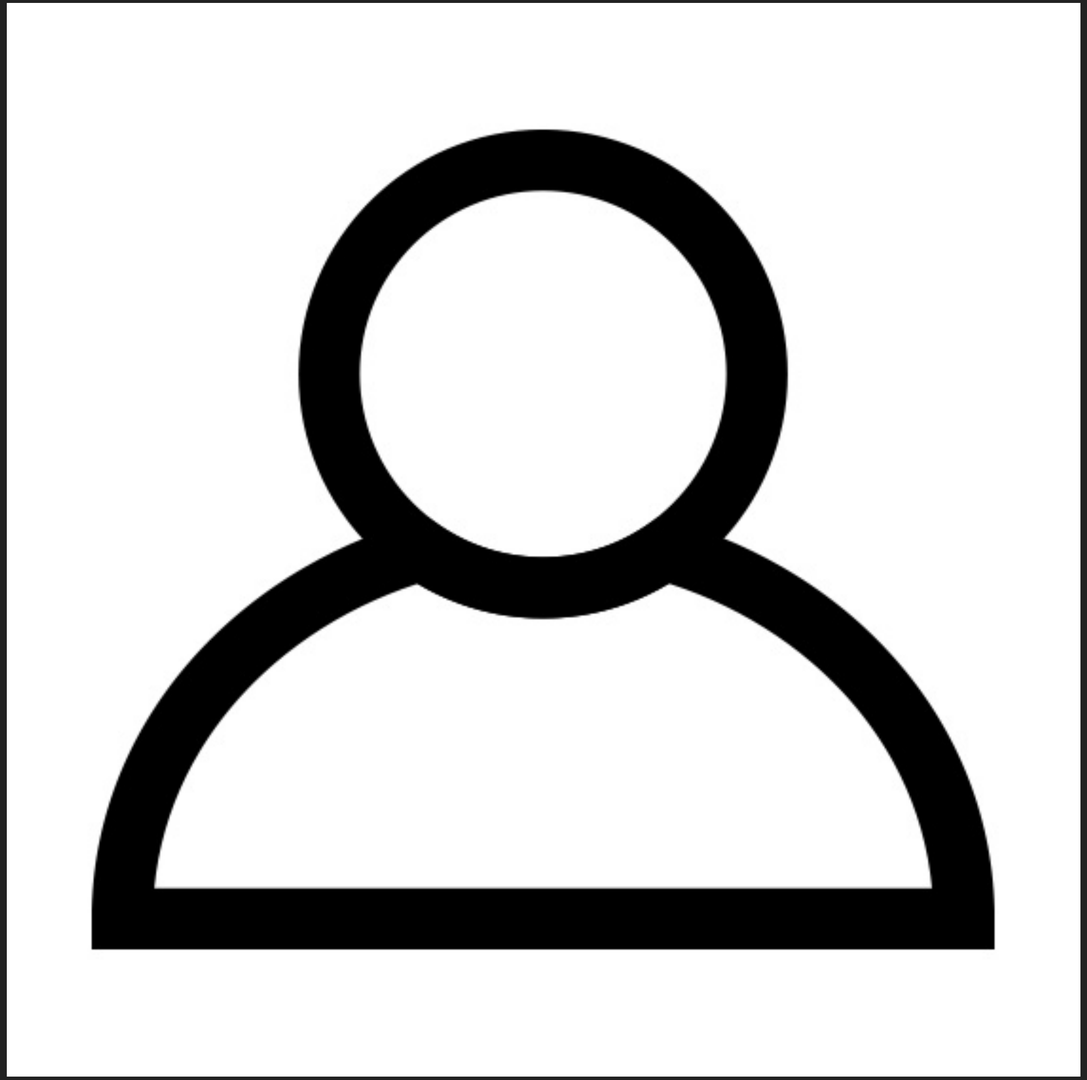
Feature engineering

- behaviors = alcohol, exercise, sleep, tobacco
- demographics = age, gender, race
- disease = asthma, diabetes, kidney, stroke, skin cancer
- measures = BMI, general, mental, physical, mobility
- risk factors = diabetes, BMI, physical activity, alcohol

Class imbalance handling

- Up/down … weight ???

Figure ##



RESULTS

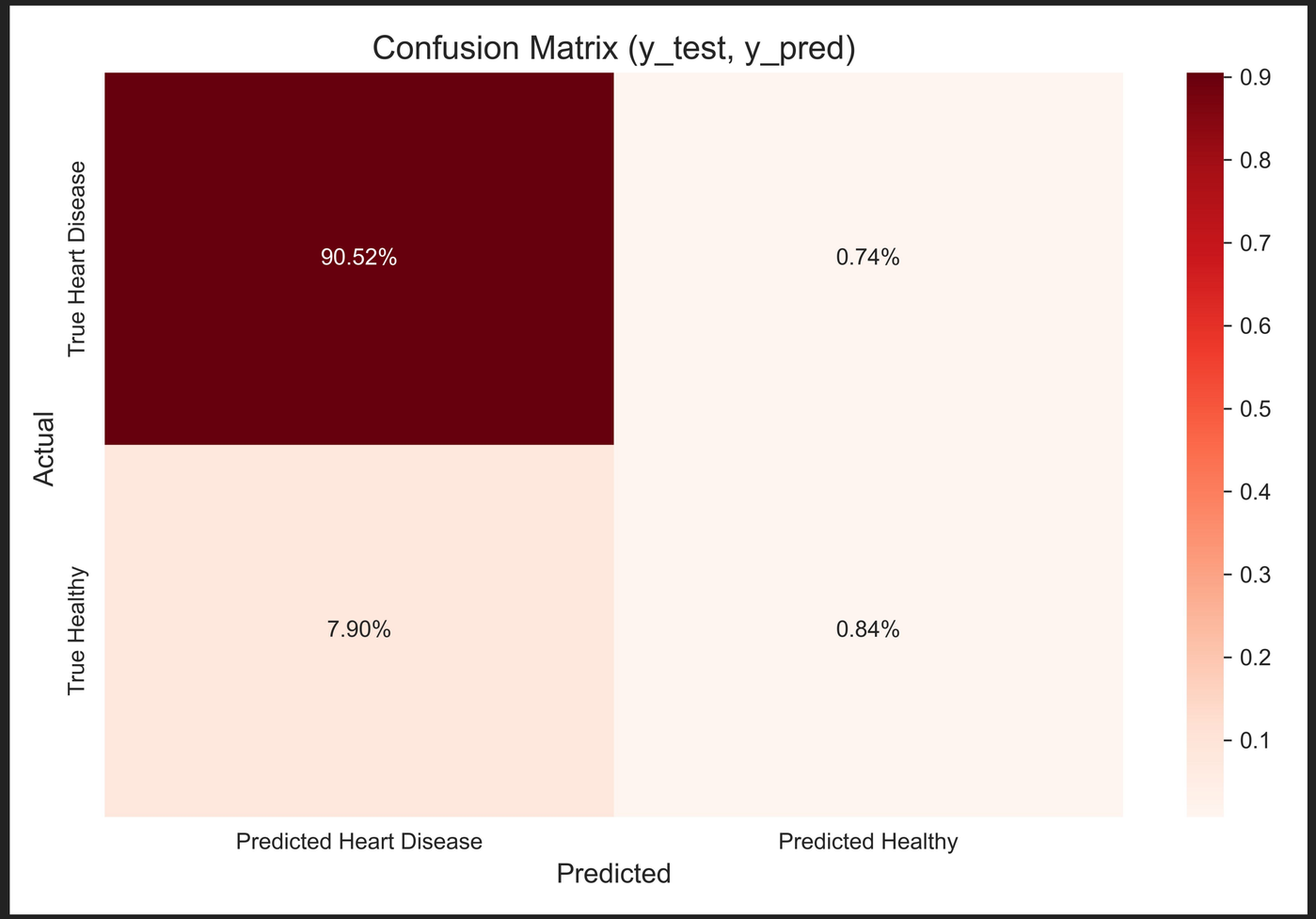
- Note: all model iterations were scored on validate split data
- GridSearchCV to tune best performing model

	Model	Recall	ROC AUC
0	Logistic regression	0.519500	0.841100
1	Decision tree (depth=2)	0.000000	0.500000
2	Decision tree (depth=4)	0.000000	0.568300
3	Random forest	0.368900	0.695500
4	XGBoost	0.084943	0.539077
5	Bernoulli NB	0.365400	0.595214
6	Gaussian NB	0.264500	0.674079
7	Multinomial NB	0.218800	0.611716
8	Hard NB Voting Classifier	0.293400	0.633538
9	Soft NB Voting Classifier	0.287000	0.637002
10	Stacked	NaN	NaN
11	Logistic regression group features	0.389500	0.797200
12	Logistic regression risk factor features	0.000000	0.646000
13	Logistic regression question groups + risk fac...	0.466800	0.799800

RESULTS

Confusion Matrix

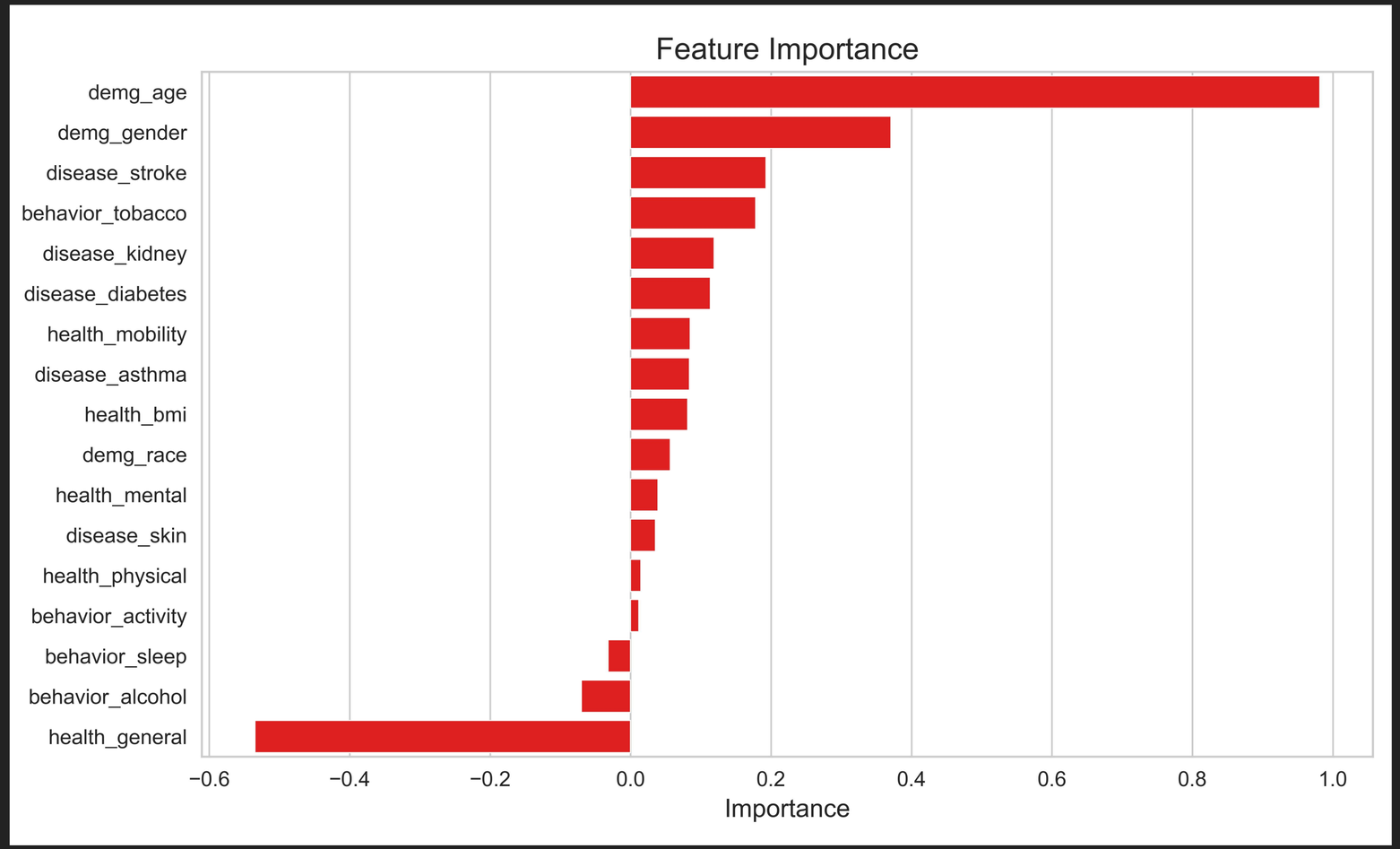
- …
- …
- …



RESULTS

Top 5 features of importance:

- ...
- ...
- ...
- ...
- ...
-



RESULTS

Predictions

- Patient #1:
- Patient #2:
- Patient #2:

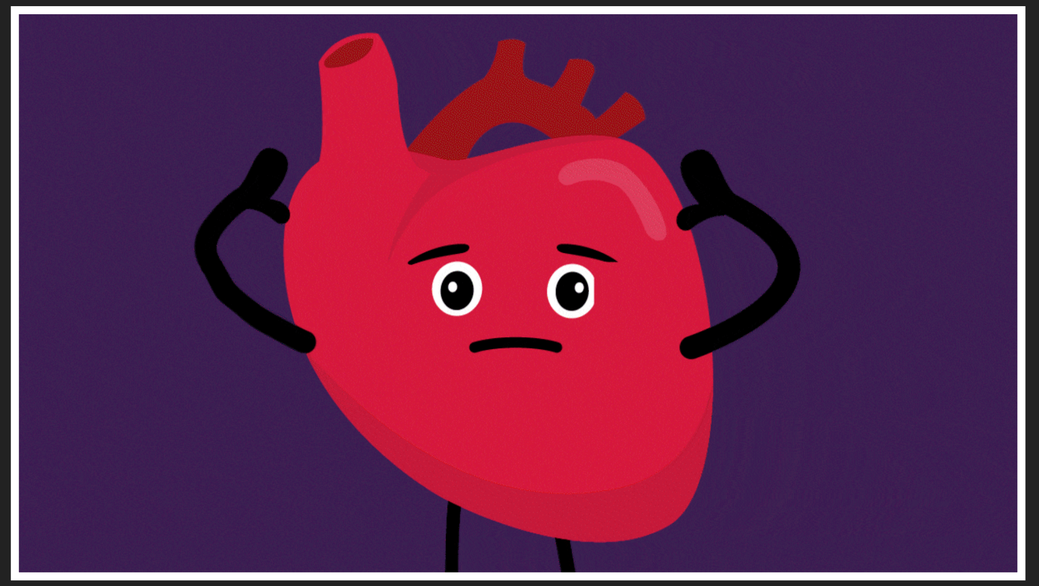
CONCLUSIONS

Insights

- …
- …
- …

Recommendations

- …
- …
- …



GIF by NHLBI #OurHearts

FUTURE WORK

- ...
- ...
- ...

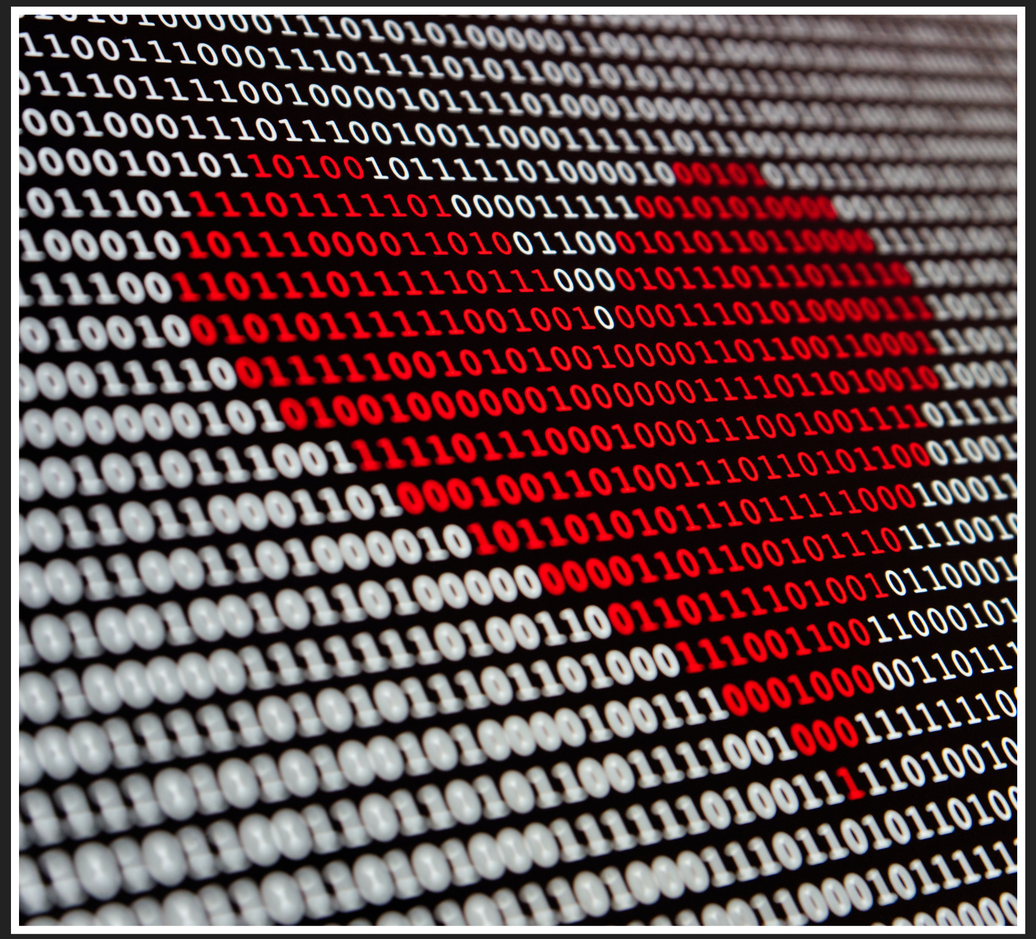


Photo by Alexander Sinn on Unsplash

APPENDIX

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

Photo by Kelly Sikkema on Unsplash

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