PREVENTING HOSPITALIZATIONS WITH MACHINE LEARNING

by Lili Beit

BUSINESS PROBLEM

- Predict which patients will be hospitalized over the next 12 months, based on demographic information and medical conditions

- Generate a list of high-risk patients, with a precision rate above 20%

- **Value-based payment systems** incentivize provider networks to deliver high-quality care at low cost

- A **predictive model** can help providers identify high-risk patients

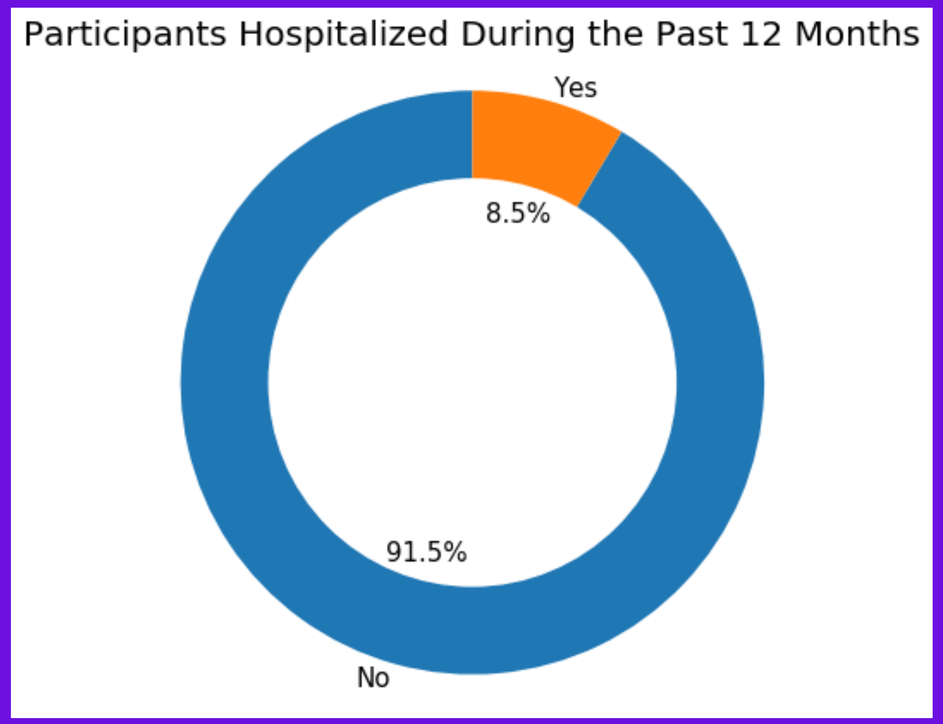
- Targeting these patients for **outreach** can prevent hospitalizations and reduce health care costs

- Provider networks that reduce costs are entitled to a share of the **savings** to insurers

DATA

NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY (NHANES)

2017-2018



SURVEY POPULATION

- Nationally representative sample of 9,000 participants, from counties across the U.S.
- Excluded participants under two years old

PREDICTORS

- Demographic information
- Medical conditions (prior to 1 year ago)
- Prescription drug use (prior to 1 year ago)

HOSPITALIZATIONS (THE TARGET)

- Only 8.5% of participants were hospitalized over the past year
- This excludes hospitalizations for childbirth

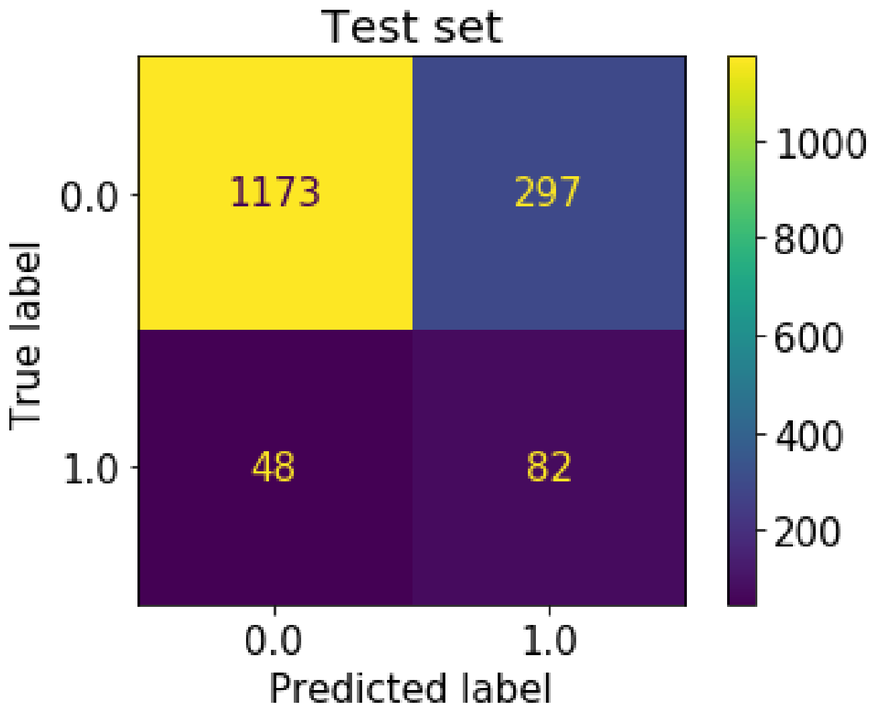
DATA GAPS AND AMBIGUITY

- Not true EHR data, relies on participant memory
- COPD
- Patients over 80

MODELING

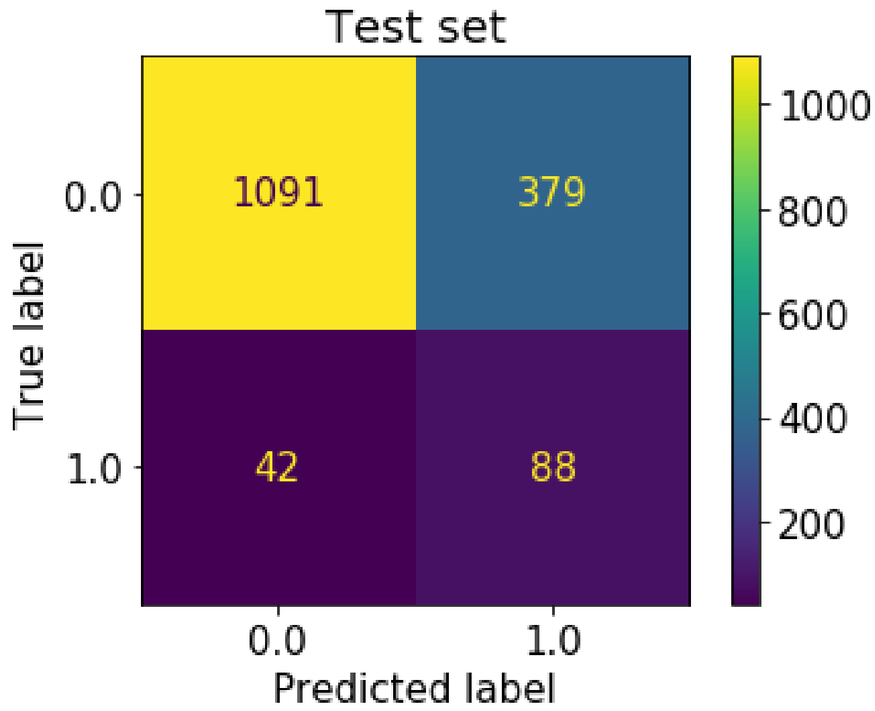
LOGISTIC REGRESSION

Recall / Sensitivity: 0.64
Precision / Specificity: 0.21

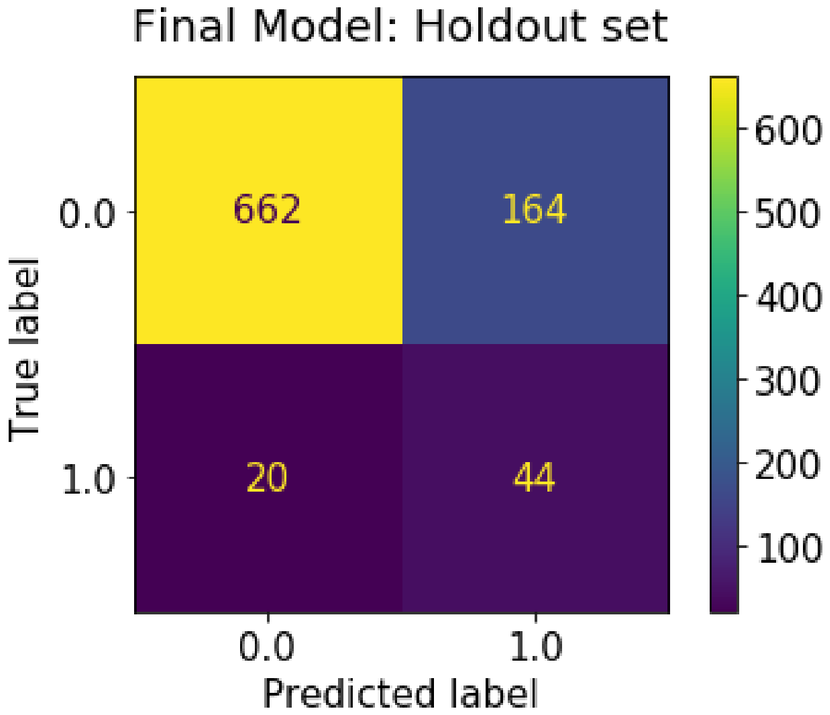


RANDOM FOREST

Recall / Sensitivity: 0.67
Precision / Specificity: 0.19



FINAL MODEL



RECALL: 0.69

PRECISION: 0.21

- LOGISTIC REGRESSION
- BALANCED CLASS WEIGHTS
- FEATURES:
  - HEART CONDITIONS
  - CIRCULATORY CONDITIONS
  - RESPIRATORY CONDITIONS
  - NUMBER OF RX DRUGS
  - ARTHRITIS
  - CANCER
  - LIVER CONDITIONS
  - AGE
  - RACE

CONCLUSIONS AND FUTURE WORK

Model effectively flags patients at high risk for hospitalization

By closely managing these patients, ACOs can prevent hospitalizations and other adverse events

Decision tree-based models may perform better with more data, and merit further study

More data points about each participant may also improve the model if the dataset were bigger

THANK YOU

QUESTIONS?

Lili Beit
Email: lilisbeit@gmail.com
github: @lilisbeit
LinkedIn: www.linkedin.com/in/lili-beit-33bb823/