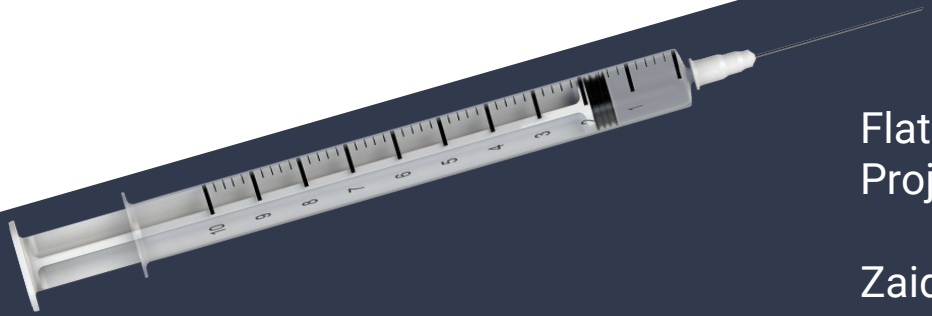
# Predicting Vaccinations Against Infectious Diseases

Using machine learning to understand
the decision to get (or not get) the shot



Flatiron Data Science
Project 3

Zaid Shoorbajee

# Outline

- Business problem
- Data understanding
- Modeling
- Model evaluation
- Recommendations

# Business Problem

How can we increase rates of vaccination against infectious diseases?

A public health agency wants to understand what factors can predict someone's decision to get vaccinated.

- Demographics?
- Socioeconomic status?
- Healthcare access?

The agency wants to use that information to reach vaccine-hesitant people.

Findings can help in the COVID-19 response and beyond.

# Data Understanding

**National 2009 H1N1 Flu Survey (CDC)**

Phone surveys about H1N1 vaccine between Oct. 2009 and June 2010.

- 26,707 respondents
- 35 features in each survey:
- Demographics, behavioral, socioeconomic, healthcare, geographic, employment

**Target: Did respondent get the H1N1 vaccine?**

About 21% of respondents got the shot.

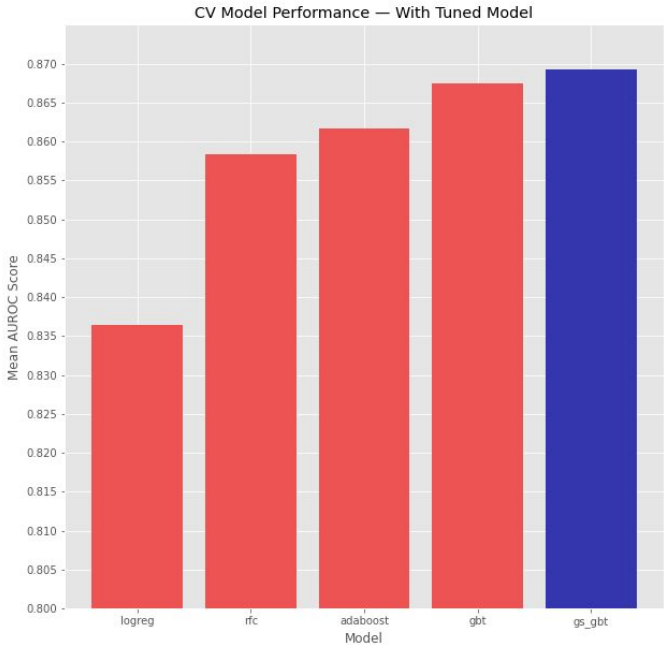
# Modeling

**Why machine learning**

- Can take in multiple features
- Identify which ones are most predictive of outcome
- Automated process

**AUROC score**

- How well can the model discriminate between positive cases and negative cases



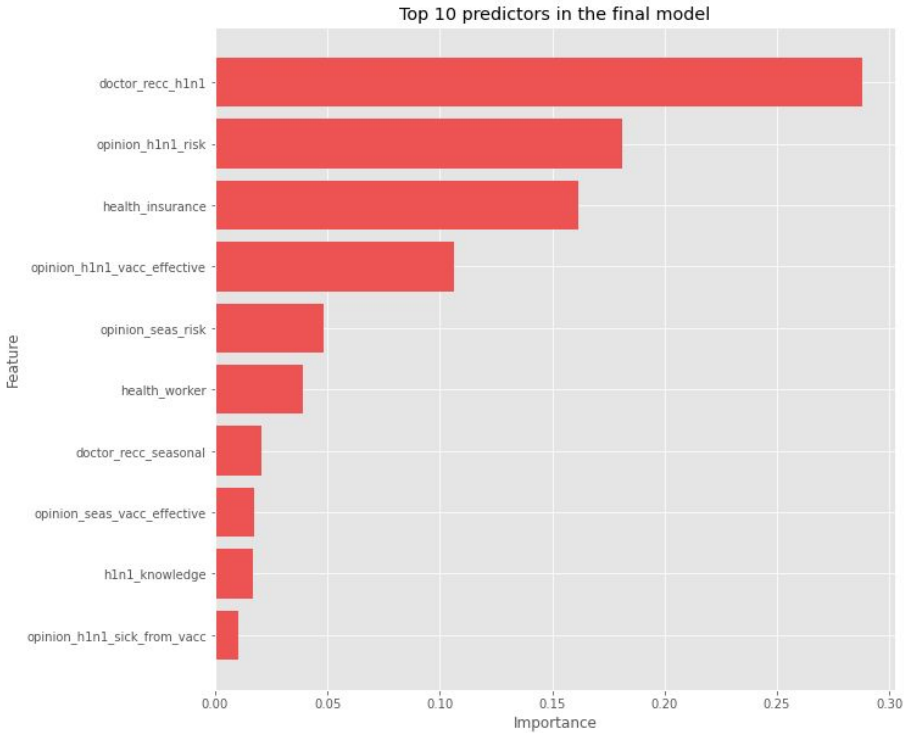
# Evaluation

**Final model:**

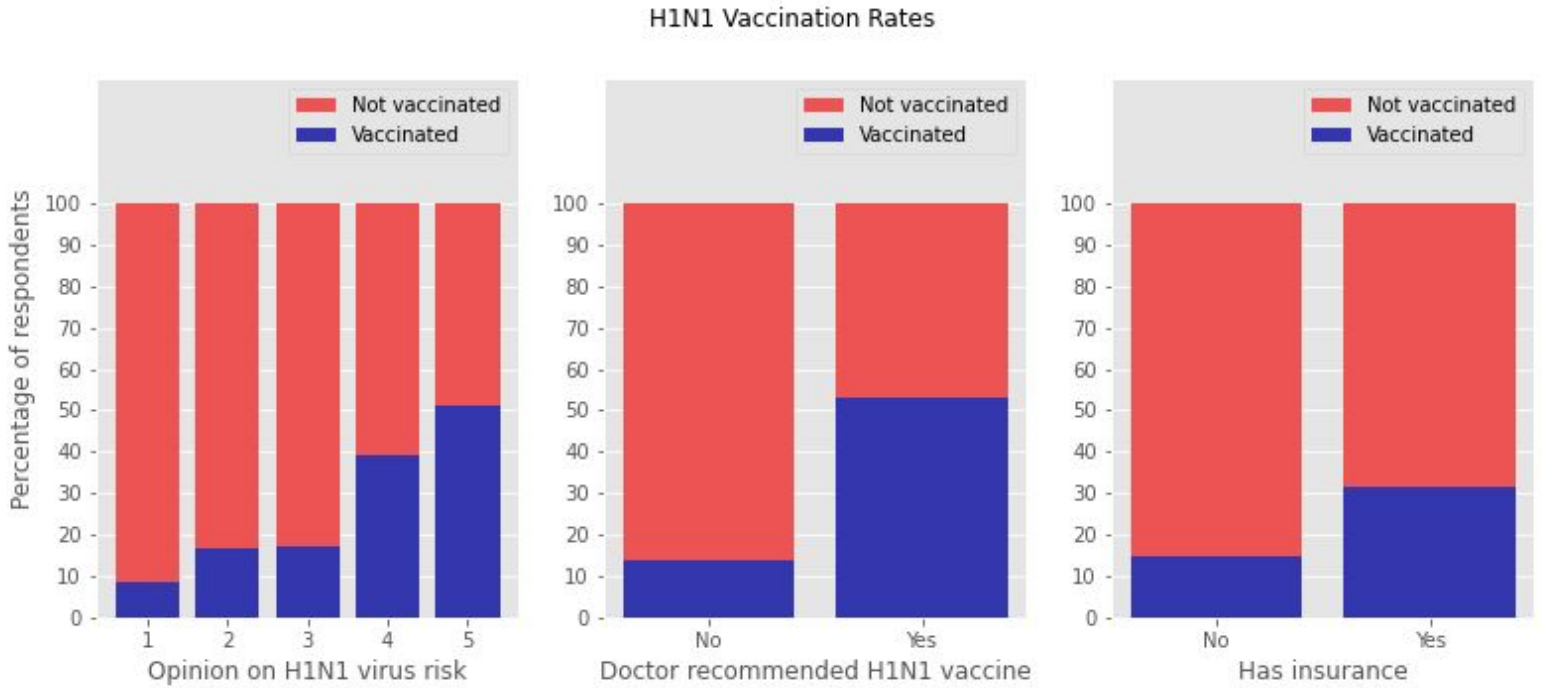
**Gradient boosted trees** algorithm

- Learns from its mistakes
- 0.8537 AUROC score

**Top predictor:** Did a doctor recommend getting the H1N1 vaccine?



# Top three predictors



# Recommendations

**Doctor's orders:** Instruct doctors and health practitioners to recommend disease-preventing vaccines.

**PSAs:** Raise public awareness about the risk the virus poses without a vaccine

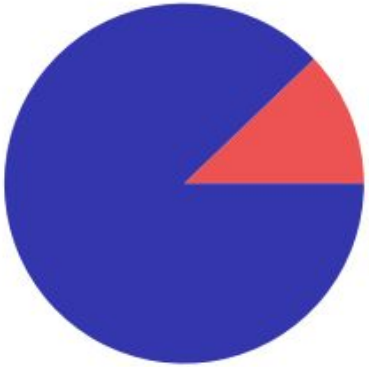
**Access:** Increase access to healthcare. Vaccine should be available to those without coverage.

Worst-case scenario

Not Vaccinated: 94.3%
Vaccinated: 5.7%

- No doctor recommendation
- Low opinion of risk (1)
- No health insurance

Best-case scenario



Not Vaccinated: 12.2%
Vaccinated: 87.8%

- Doctor recommended vaccine
- High opinion of risk (5)
- Has health insurance

## Top three predictors

# Next steps

**H1N1 data is from over 10 years ago**

COVID-19 response has been different and more intense.

Would this analysis generalize to information about COVID-19?

Find similar data about COVID-19 and compare results.

Findings can inform response to other public health crises.

# Thank you

Email: zshoorbajee@gmail.com

GitHub: @zshoorbajee

LinkedIn:
linkedin.com/in/zshoorbajee/