

# Flu Vaccine Intake Classification

Aysu Erdemir  
November, 2022



**Stakeholder:** CDC – Center for Disease Control and Prevention

---

**Business  
Problem:**

- CDC is up for a vaccine campaign to maximize vaccine intake for the upcoming flu season.
- They want to know what factors are most important in predicting whether a person would take the vaccine or not.

# Data:



National 2009 H1N1 Flu Phone Survey from 26707 participants, which contains information about:

- ✓ whether people had received the **seasonal flu vaccine** or not, AND
- ✓ their demographic background,
- ✓ their opinions,
- ✓ their health behaviors

---

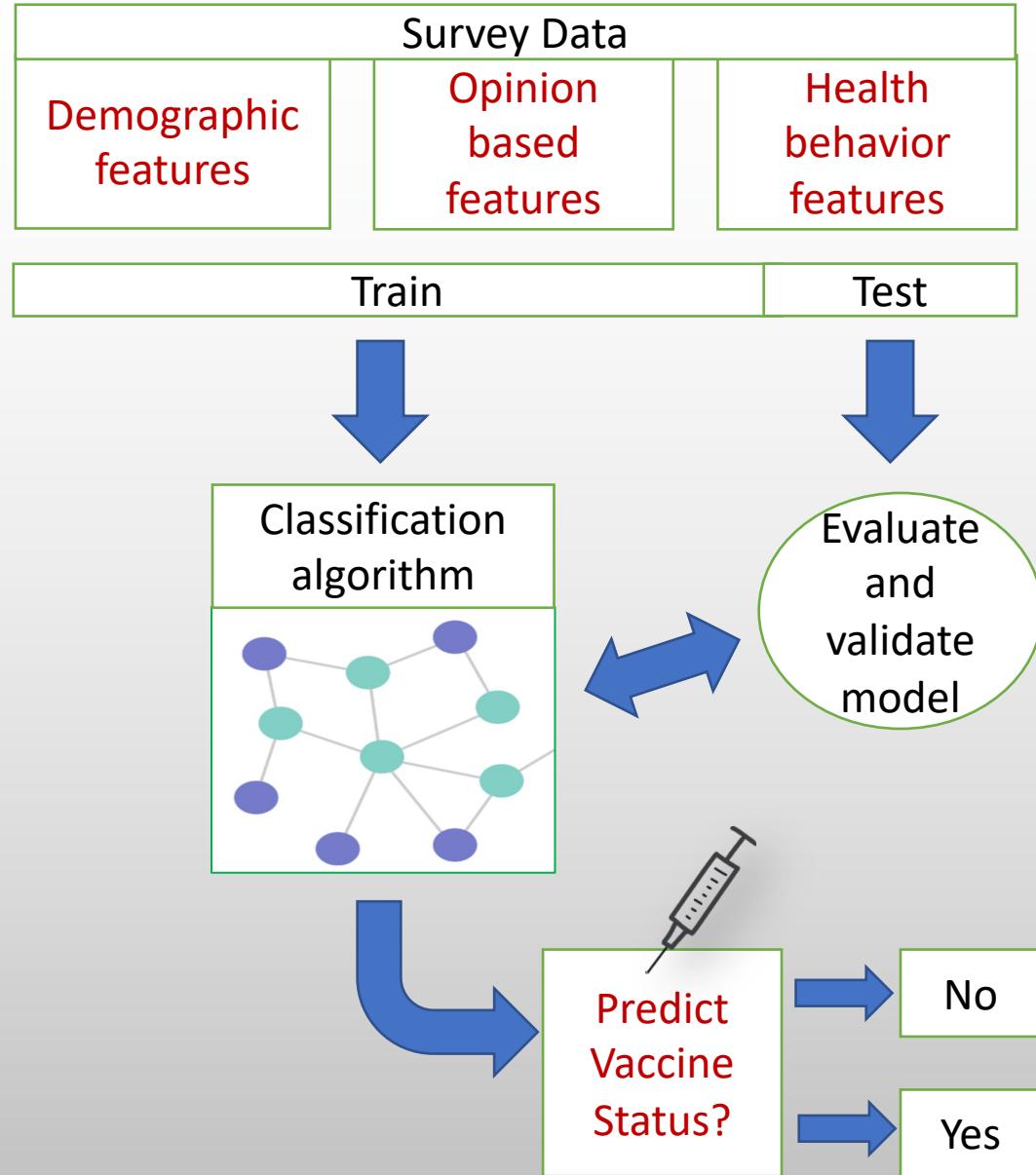
# Goal:



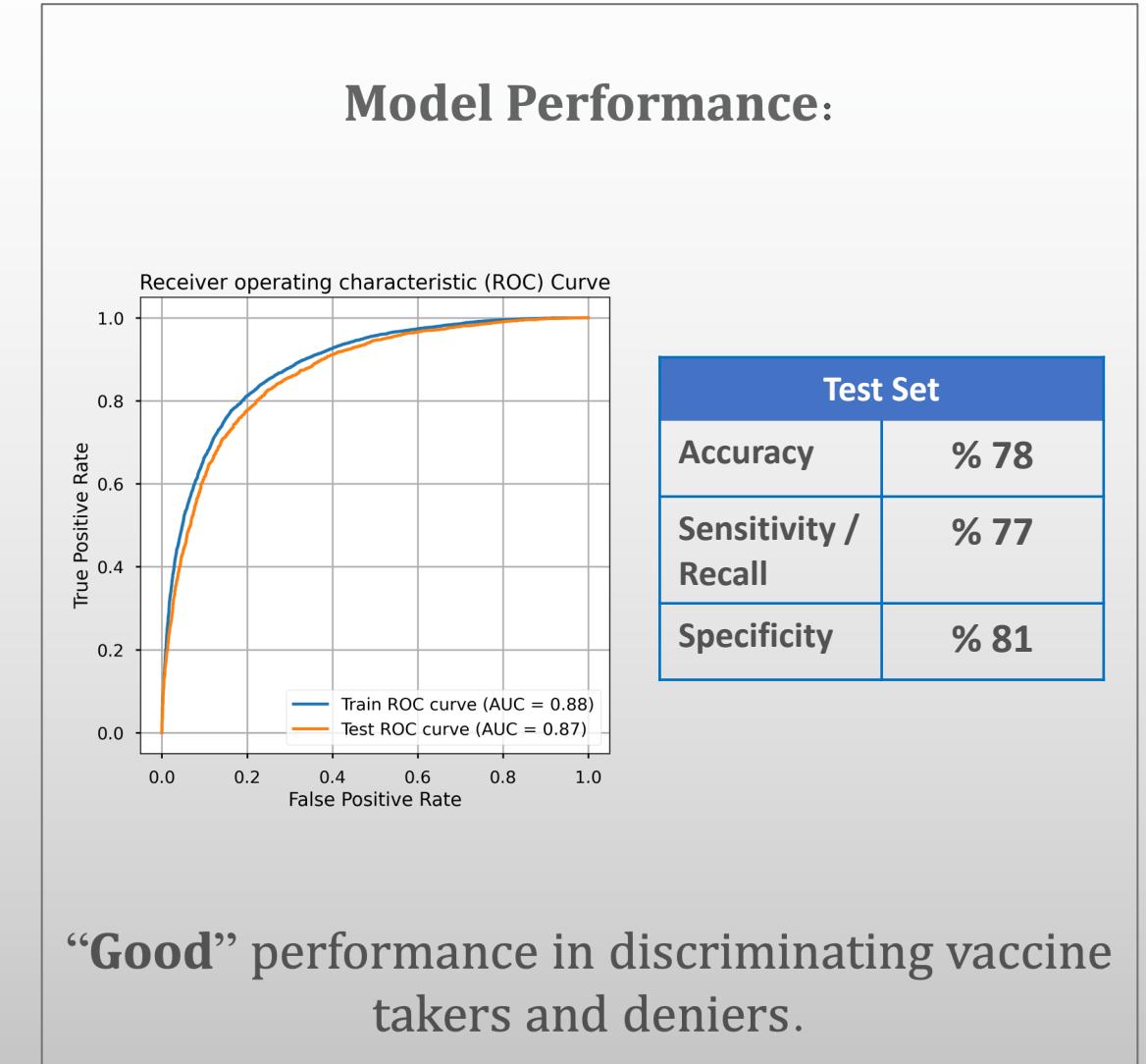
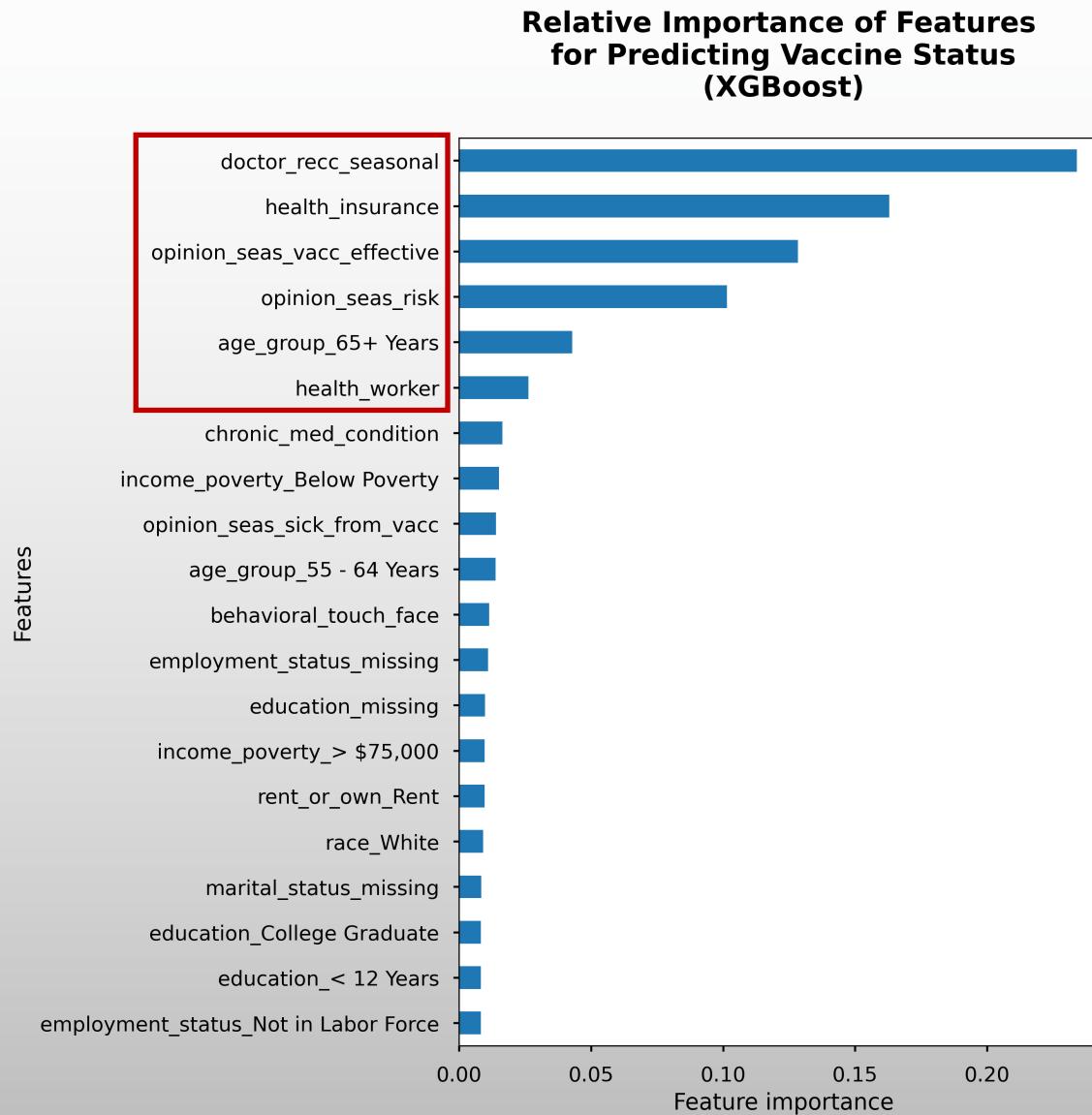
- Build a **classifier** that predicts vaccination status using available information.
- Extract the likelihood of getting vaccinated for each important feature to be able to make recommendations.

# Modeling:

- Clean and preprocess the data.
- Build, tune and validate several types of machine learning classifiers:
  - ✓ Logistic Regression
  - ✓ Decision Tree
  - ✓ Random Forest
  - ✓ XGradient Boosted
  - ✓ Stacking Classifier
- Sort most important features for predicting vaccine outcome.
- Make specific recommendations.

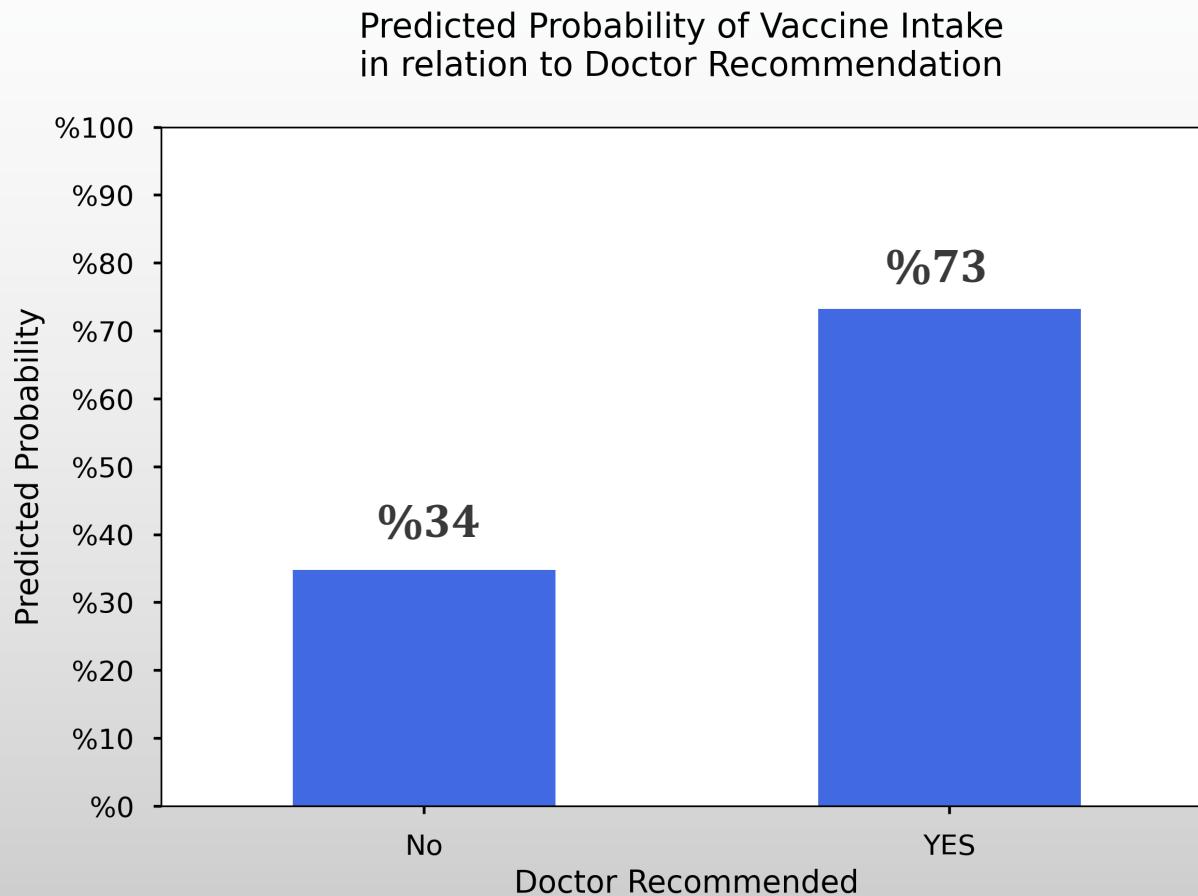


# Results from top model:



*The most important predictive feature:*

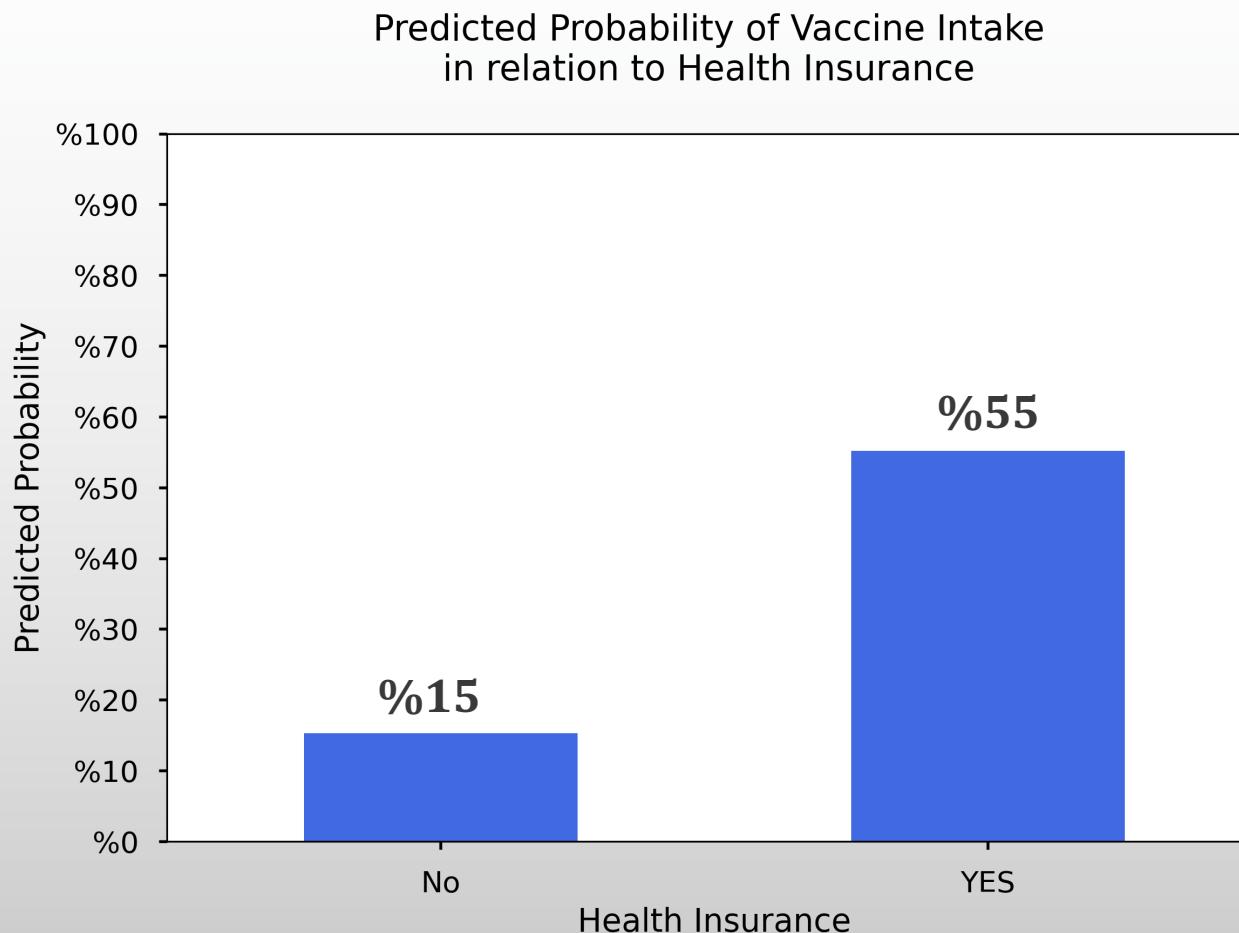
# Does your doctor recommend getting flu vaccine?



- People are **%73** likely to receive the vaccine if it was recommended by their doctor.
- People are only **%34** likely to receive the vaccine if it was **NOT** recommended by their doctor.

Target physicians by educating them on the importance of vaccination & recommending it to their patients!

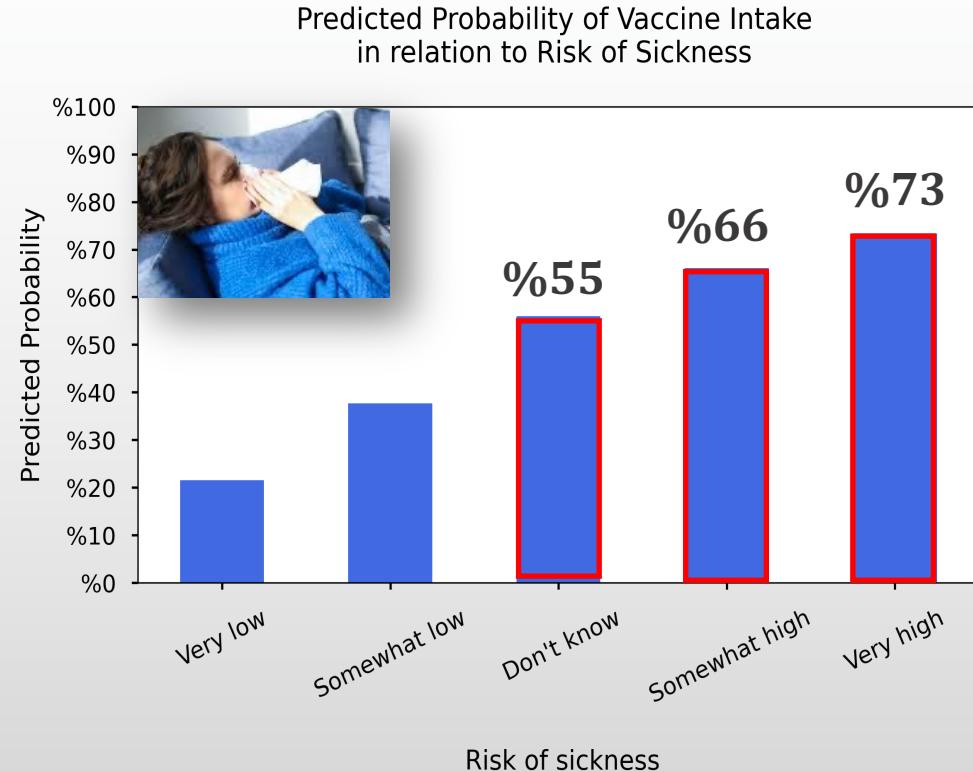
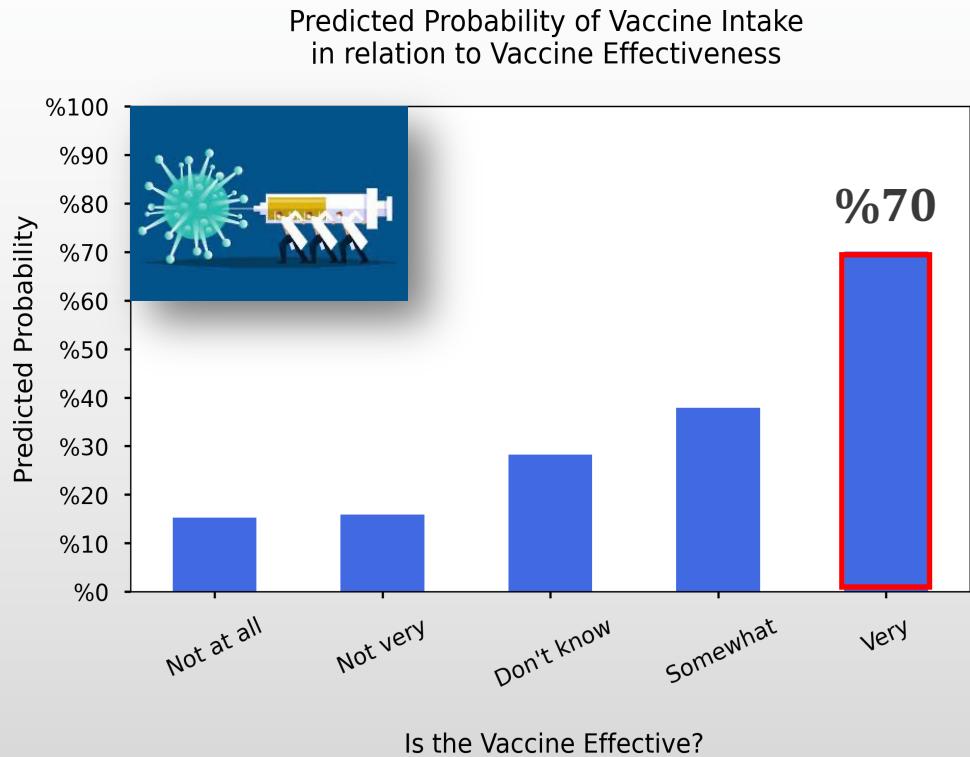
# Do you have health insurance?



- People with health insurance are substantially more likely to get the vaccine (%55 compared to %15).

Target uninsured populations in the campaign, but better yet work on universal health coverage for all individuals and communities.

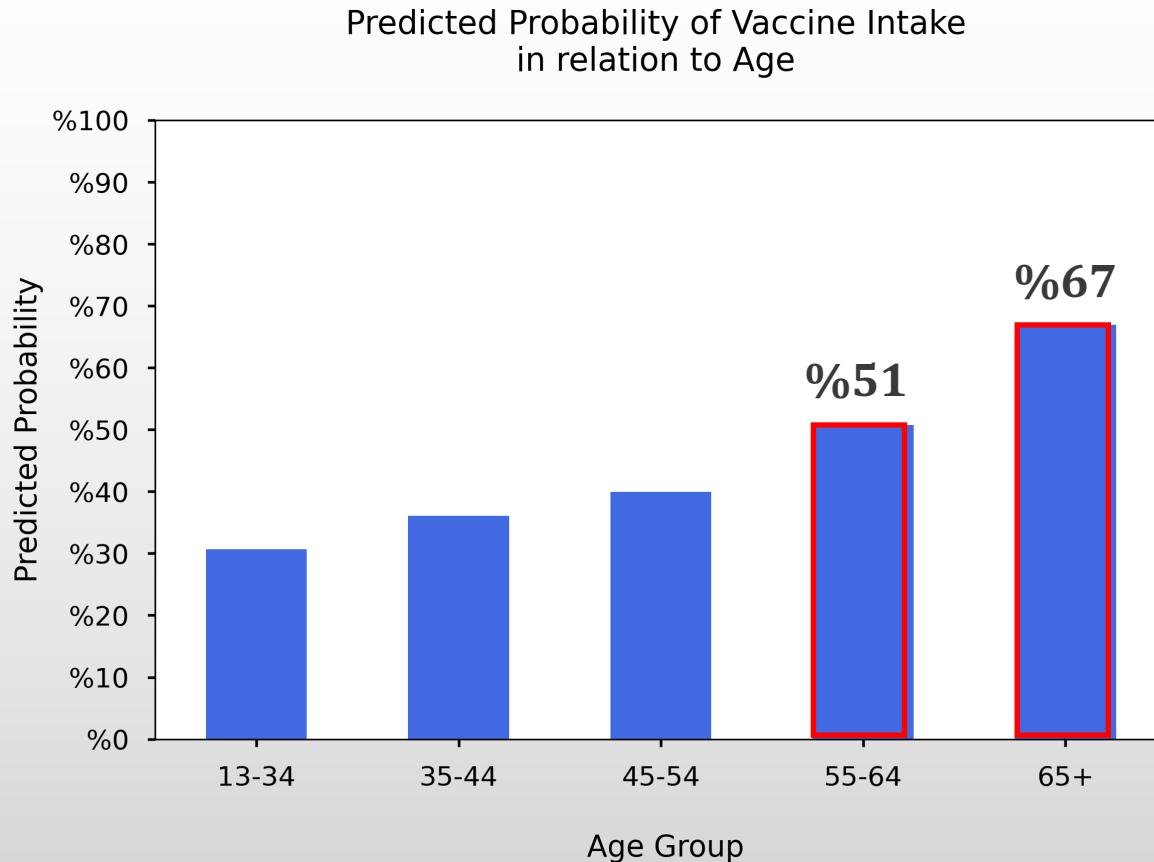
# Do you think flu vaccine is effective? Are you worried about getting sick if not vaccinated?



- Only people who rated the vaccine as very effective were more likely to have gotten the vaccine (%70)
- As people's worry about getting sick increased their likelihood to get the vaccine also increased.

**Focus your campaign on informing the people about the effectiveness and safety of the vaccine or their risk of falling ill and developing complications if not vaccinated.**

# How old are you?



- **65+ year old** people are much more likely to get the flu vaccine (**%67**).
- **55-64 year olds** are slightly more likely to receive the vaccine (**%51**).

As a priority keep focusing your campaign on older age groups, because they are at more risk of developing flu-related complications compared to younger age groups. There is still room for progress!

Also target younger people as a key demographic population since their vaccination rates are much lower.

# Conclusions

You are more likely to get the vaccine if:

- ✓ your doctor recommends the vaccine
- ✓ you have health insurance
- ✓ you think the vaccine is effective
- ✓ you believe you can get sick from flu
- ✓ you are +65 years old

# Recommendations

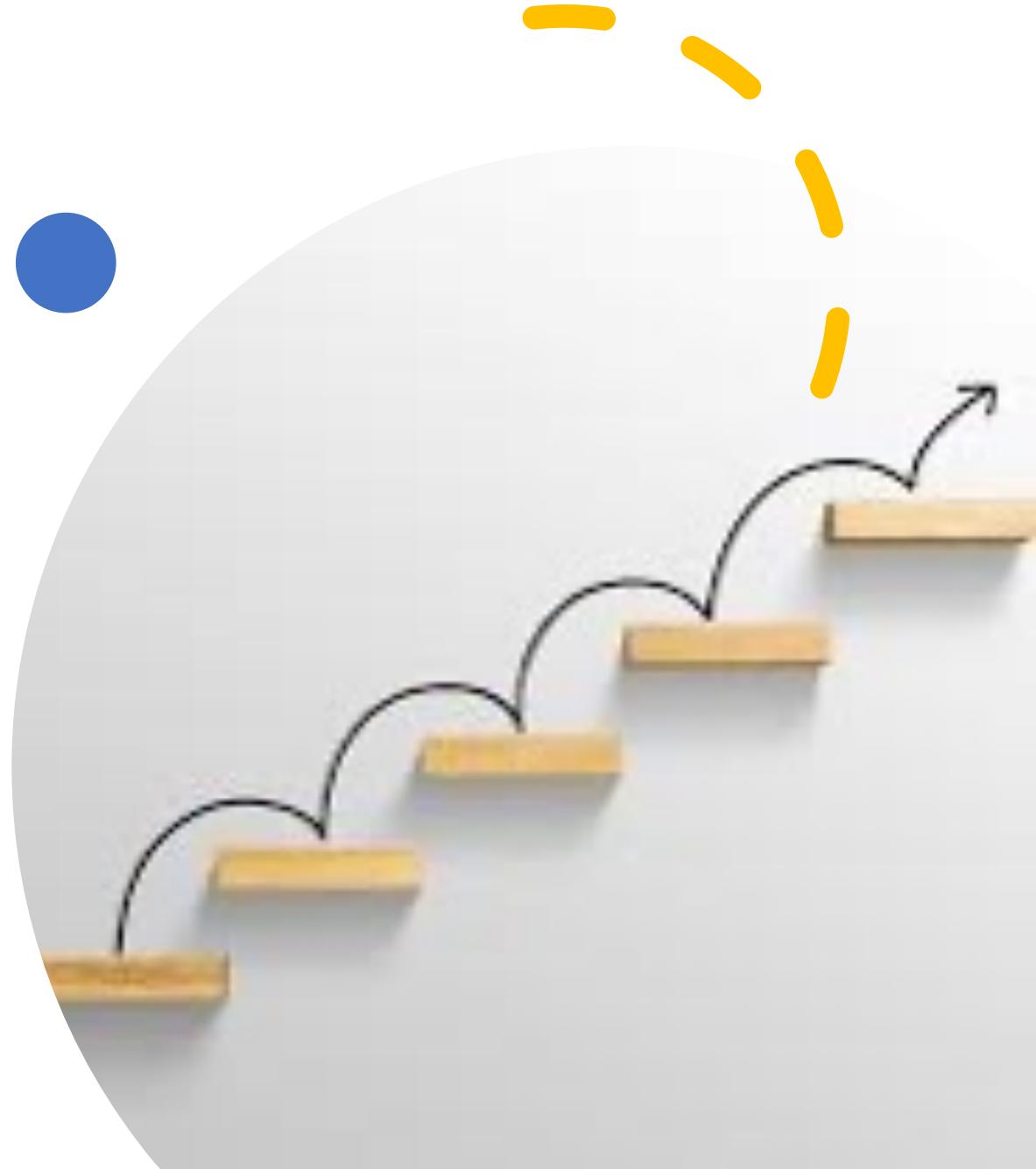
For maximizing vaccine intake:



- Target **physicians** by educating them on the importance of vaccination & recommending it to their patients.
- Target **uninsured** populations, but better yet work on universal health coverage.
- Inform the people about the **effectiveness** and safety of the vaccine and their **risk** of falling ill and developing complications if not vaccinated.
- Keep focusing on **older** age groups as they are at more risk of developing complications from flu. But also target **younger** people to maximize the benefits of herd immunity since they are much less likely to receive the vaccine.

# Limitations and Improvements

- Encrypted **employment industry, employment occupation, and geographical region** info, hard to make any specific suggestions based on these features.
- Results on **health insurance** are not very reliable due to having %40 missing data which was encoded using predictive modeling. Emphasis needs to be given to this variable next time the survey is conducted.
- More recent data needs to be collected after the Covid-19 **pandemic** since the pandemic might have altered people's attitude towards flu vaccine as well.





Email: erdemiraysu@gmail.com

GitHub: @erdemiraysu

LinkedIn: linkedin.com/in/aylsrudemir/