A PREDICTIVE MODELING APPROACH TO UNDERSTAND AND ADDRESS H1N1 VACCINE HESITANCY

Presentation By: Irene Kibengo

OVERVIEW

Background: H1N1 Pandemic & Vaccination

The 2009 H1N1 pandemic caused up to 575,000 deaths globally in its first year. A vaccine was introduced in October 2009 in the U.S. The National H1N1 Flu Survey was launched to understand public response.

Survey data captured:

- Vaccination status
- Demographics & economic info
- Beliefs on vaccine safety/effectiveness
- Preventive behaviors

IDENTIFIED PROBLEM

Public health agencies need to boost vaccine uptake during pandemics.
This project analyzes 2009 H1N1 survey data to uncover patterns in vaccination behavior.
By predicting who is likely to get vaccinated, we aim to support targeted, data-driven strategies that address vaccine hesitancy and improve future campaign effectiveness.

PROJECT GOAL

To uncover patterns in H1N1 vaccination behavior using survey data and predictive modeling.

OBJECTIVES

- Identify key factors influencing vaccine acceptance.
- Build a model to predict likelihood of vaccination.
- Provide actionable insights to support targeted public health strategies.
- Help address vaccine hesitancy in future pandemic responses.

DATA SOURCE

National 2009 H1N1 Flu Survey (NHFS), USA — collected post-pandemic to assess vaccine uptake, behaviors, and opinions.

The Dataset includes the following:

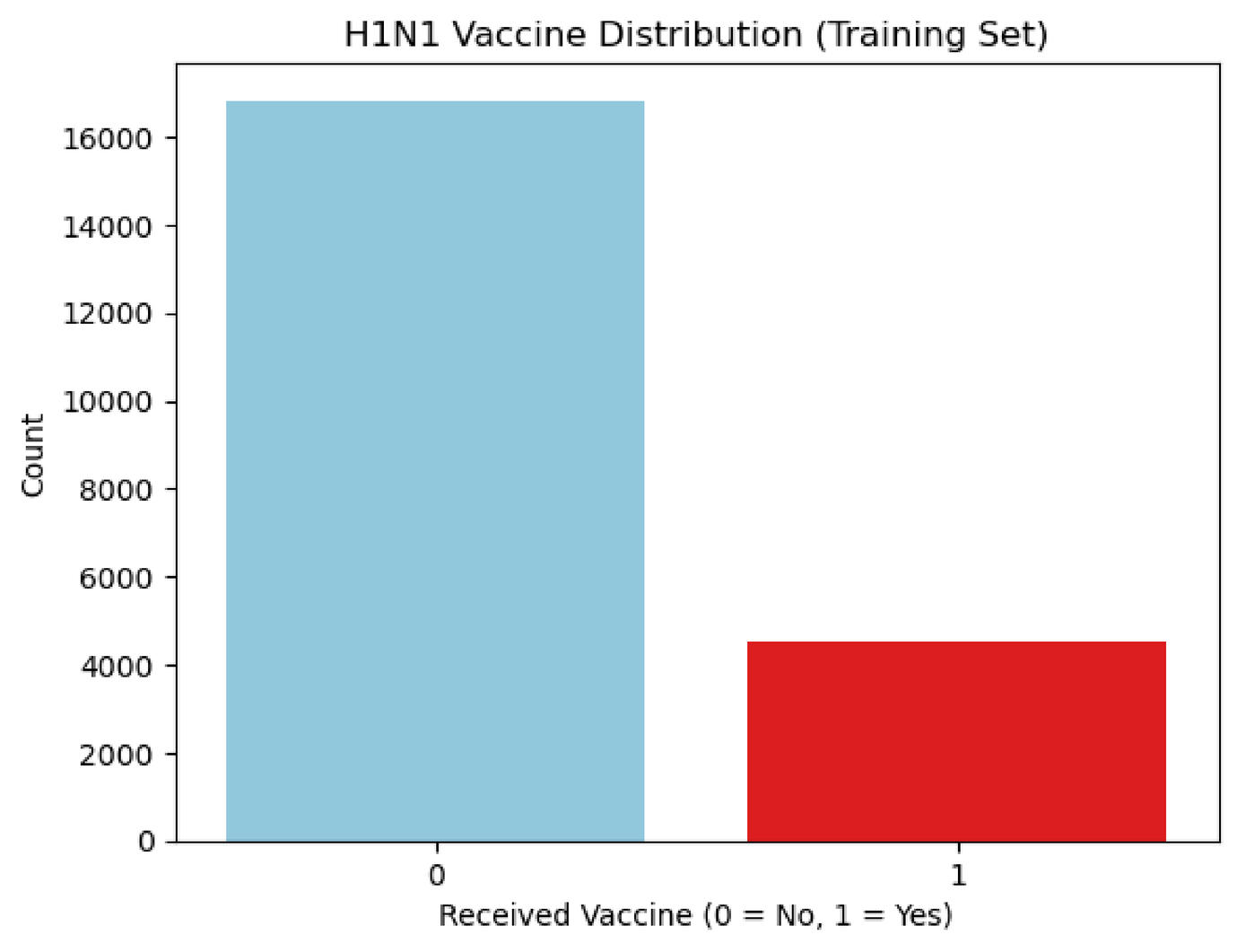
1.Demographic Features
2.Behavioural Features
3.Opinion- Based Features
4.Target Variable- H1N1 Vaccine

Modeling Approach

To predict who received the H1N1 vaccine, I tested three models:

1. Logistic Regression - A simple, interpretable model that provides a strong baseline for understanding vaccine uptake patterns.
2. Balanced Logistic Regression- Adjusted to handle class imbalance—helps better identify vaccinated individuals, which is important for targeted outreach.
3. Random Forest- A more advanced model that captures complex patterns and improves performance, especially when dealing with many unvaccinated individuals.

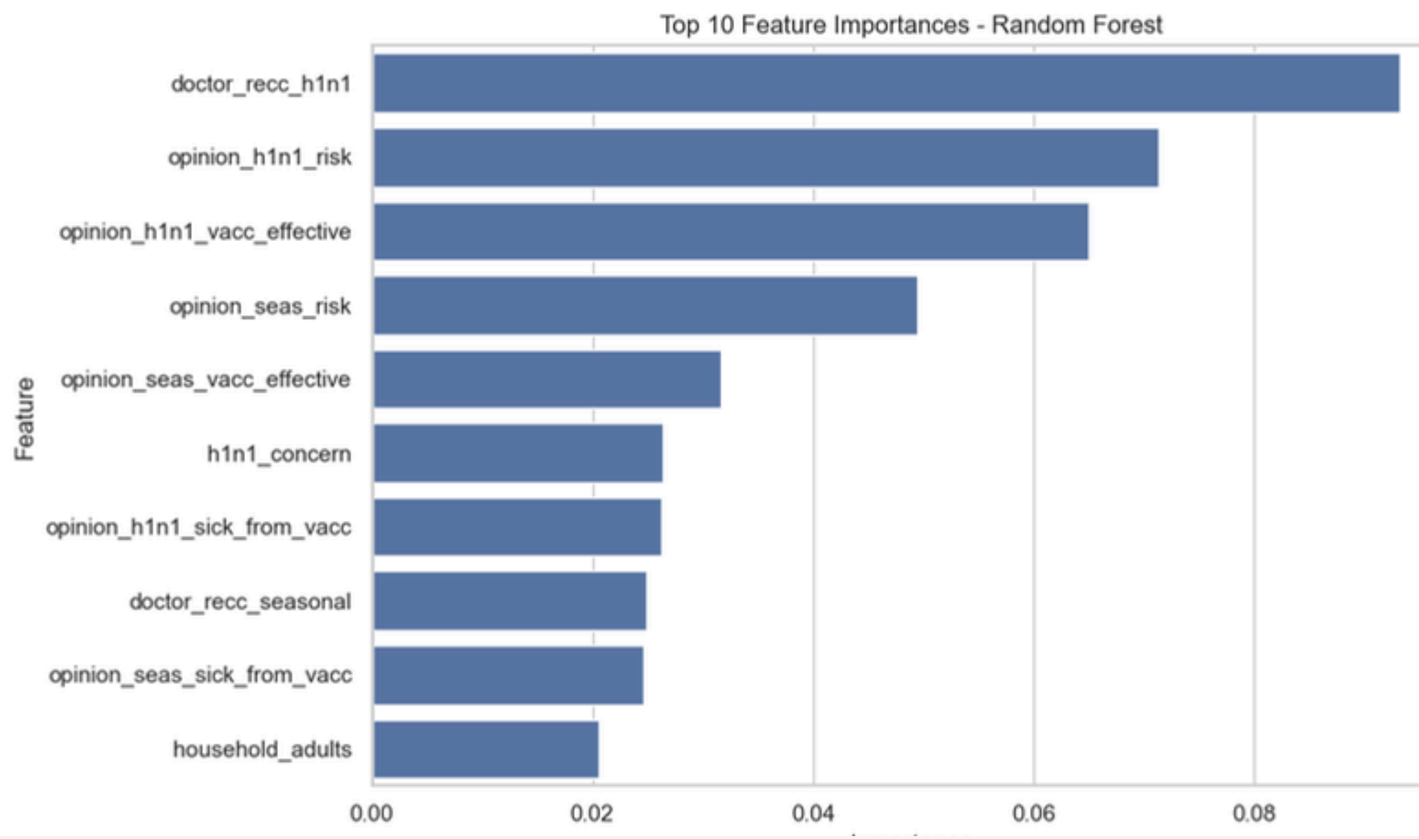
TARGET DISTRIBUTION



The bar chart illustrates the distribution of the target variable:

- Label 0 represents individuals who did not receive the H1N1 vaccine.
- Label 1 represents individuals who did receive the vaccine.

TOP 10 FEATURE IMPORTANCE



The strongest predictors of H1N1 vaccination are:

- Doctor's recommendation (doctor_recc_h1n1): People are much more likely to get vaccinated when their doctor recommends it.

This highlights that trust in medical advice and belief in vaccine efficacy are key drivers of vaccination behavior.

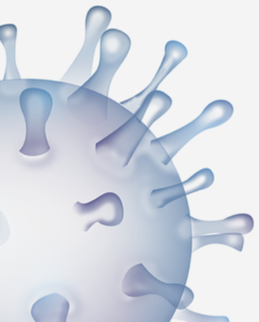
Negative influences include:

- Fear of side effects (opinion_h1n1_sick_from_vacc, opinion_seas_sick_from_vacc): Individuals who think the vaccine might make them sick are less likely to get vaccinated.

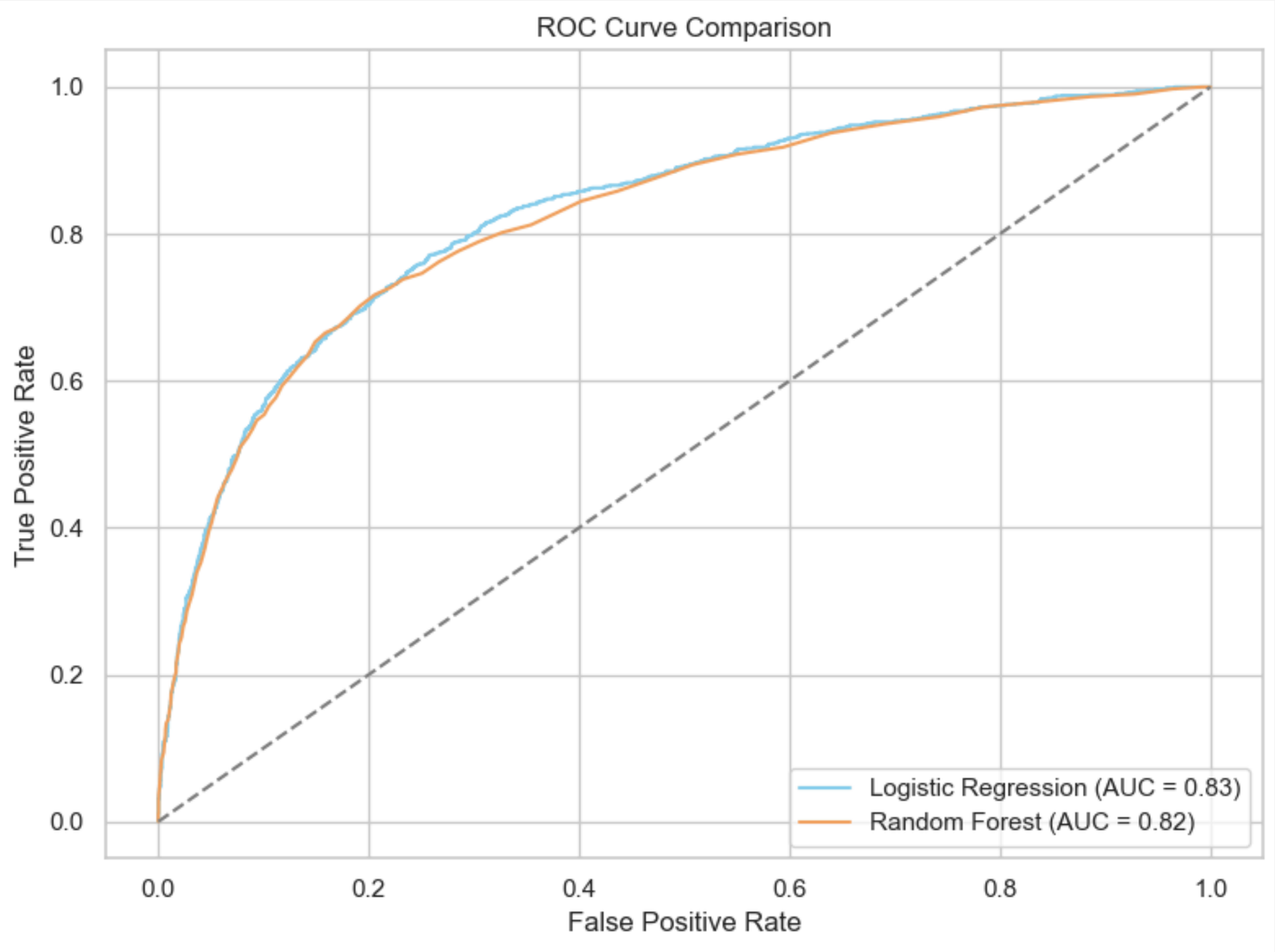
EVALUATION

Model Performance Summary

1. Logistic Regression (Balanced)
   - This model was the best at identifying people who got the vaccine, successfully finding about 72% of them.
   - It provides a strong overall balance between finding and correctly labeling vaccinated individuals, making it reliable for public health decisions.
2. Random Forest
   - This model had slightly better overall accuracy, meaning it got more total answers right.
   - However, it was not as good at identifying vaccinated people, missing many of them.
   - When it did predict someone got the vaccine, it was usually correct — but it missed more than it found.



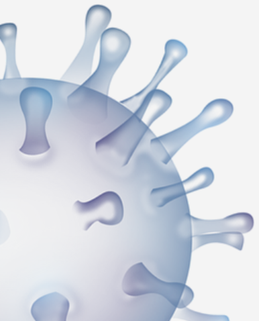
ROC CURVE COMPARISON



Logistic Regression slightly outperformed Random Forest in identifying vaccinated individuals across thresholds.

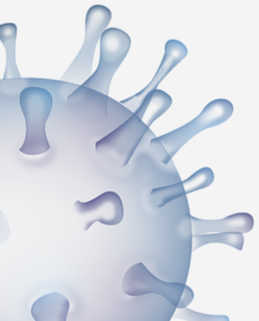
CONCLUSION

- Balanced Logistic Regression is the most effective model for identifying vaccinated individuals, with a strong recall of 72%.
- It offers a good balance between recall, precision, and AUC, making it suitable for public health decision-making.
- Random Forest, while more accurate overall, struggles to detect vaccinated individuals due to low recall.



RECOMMENDATIONS

1. Use Logistic Regression (Balanced) for decision-making:
- High recall and F1-score — ideal for identifying vaccinated individuals
- Transparent and easy to interpret
- Supports targeted public health outreach strategies
2. Use Random Forest for exploration:
- Valuable for analyzing feature importance
- Enhances understanding through deeper insight



WHAT THIS MEANS TO PUBLIC HEALTH TEAMS

- Target: Younger adults and individuals with low trust in vaccines or no doctor recommendation.
- Act: Use the model to guide outreach and deliver personalized, trust-building messages.
- Plan Ahead: Apply the model to future campaigns, monitor its performance, and improve it with new data.

THANK YOU

Email: irenekibengo@gmail.com
Github: @irene-coder-dot