H1N1 FLU AND SEASONAL FLU VACCINES

OVERVIEW/INTRODUCTION

- As the world struggles to vaccinate the global population against COVID-19, an understanding of how people's backgrounds, opinions, and health behaviors are related to their personal vaccination patterns can provide guidance for future public health efforts. Your audience could be someone guiding those public health efforts
- Beginning in spring 2009, a pandemic caused by the H1N1 influenza virus, colloquially named "swine flu," swept across the world approximating 151,000 and 575,000 deaths
- Vaccines provide immunization for individuals, and enough immunization in a community can further reduce the spread of diseases through "herd immunity.

BUSINESS UNDERSTANDING: H1N1 AND SEASONAL FLU VACCINES

- H1NI is used to prevent infection caused by the influenza A virus and Seasonal Flu vaccines protect against the four influenza viruses.
- As of the launch of vaccine competition, vaccines for the COVID-19 virus are still under development and not yet available
- A National Flu survey was conducted by United States in 2009, survey asked respondents whether they had received the H1N1 and seasonal flu vaccines the questions covered their social, economic, and demographic background, opinions on risks of illness and vaccine effectiveness, and behaviors towards mitigating transmission
- The potential target is whether the respondents received H1N1 flu vaccine.

PROBLEM STATEMENT

- Using data collected in the National 2009 H1N1 Flu Survey, the problem is to predict whether people got H1N1 and seasonal flu vaccine.
- The goal is to predict how likely individuals are to receive their H1N1 and seasonal flu vaccines. The potential target is whether the respondents received H1N1 flu vaccine.

OBJECTIVES

- 1. To predict whether people got H1N1 and seasonal flu vaccines using information they shared about their backgrounds, opinions, and health behaviors?
- 2. To identify whether the survey respondent received H1N1 flu vaccine.
- 3. To evaluate the performance of the model/respondents ()

DATA SOURCE

The source of data is using data collected in the National 2009 H1N1 Flu Survey in United States.

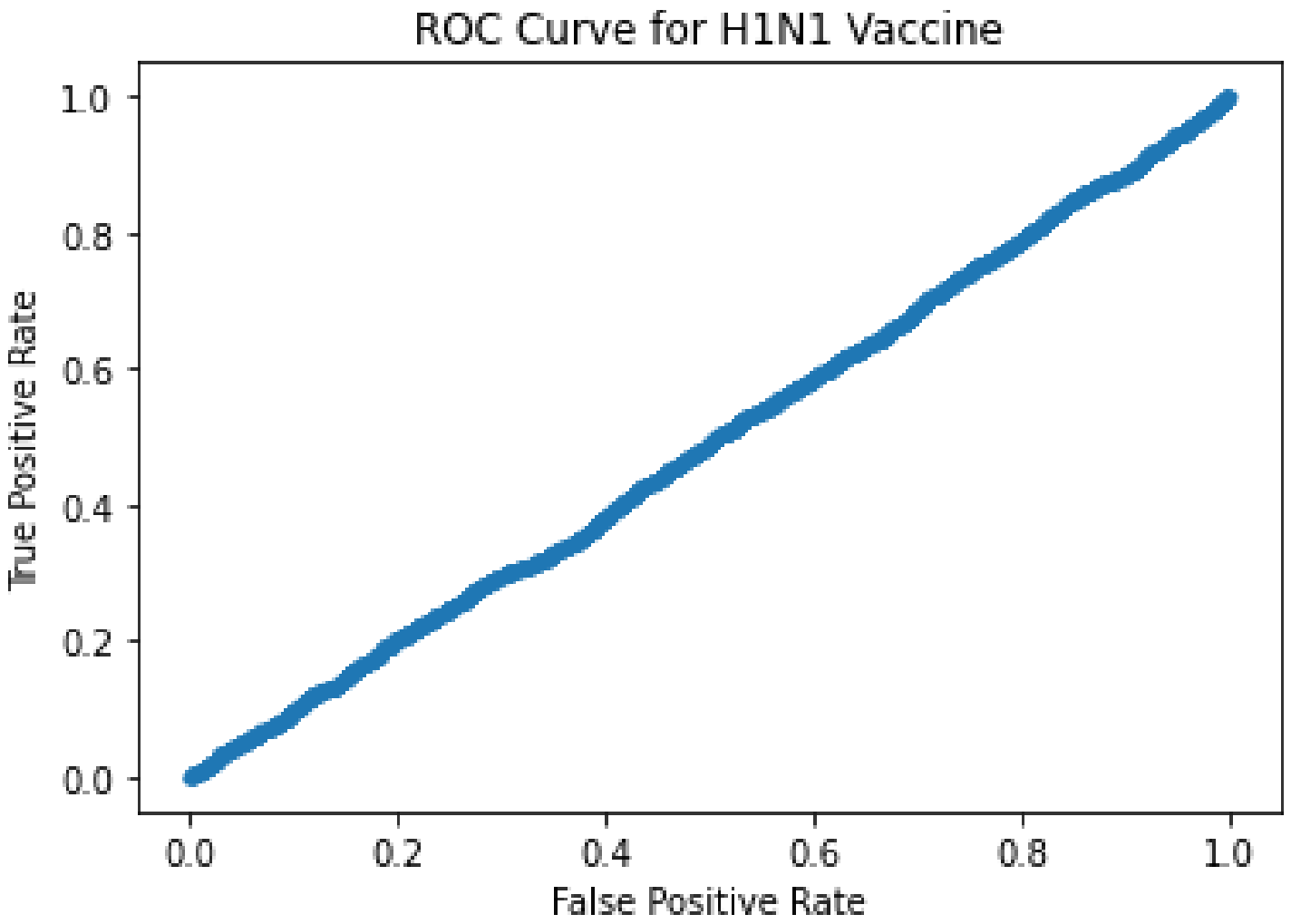
Data preparation and cleaning

- After importing the data (**training_set_labels**) we were able to have the table containing both vaccines. The table indicates the number of respondents who received and did not receive th7 vaccines.
- The total number of respondents were from 0 – 26708 for training sets and from 26708 – 5674 for test sets.
- Data preprocessing was done to identify and separate the target variables and the features.
- Categorical values were transformed to numerical data
- Train test split was done to identify X and y variables.

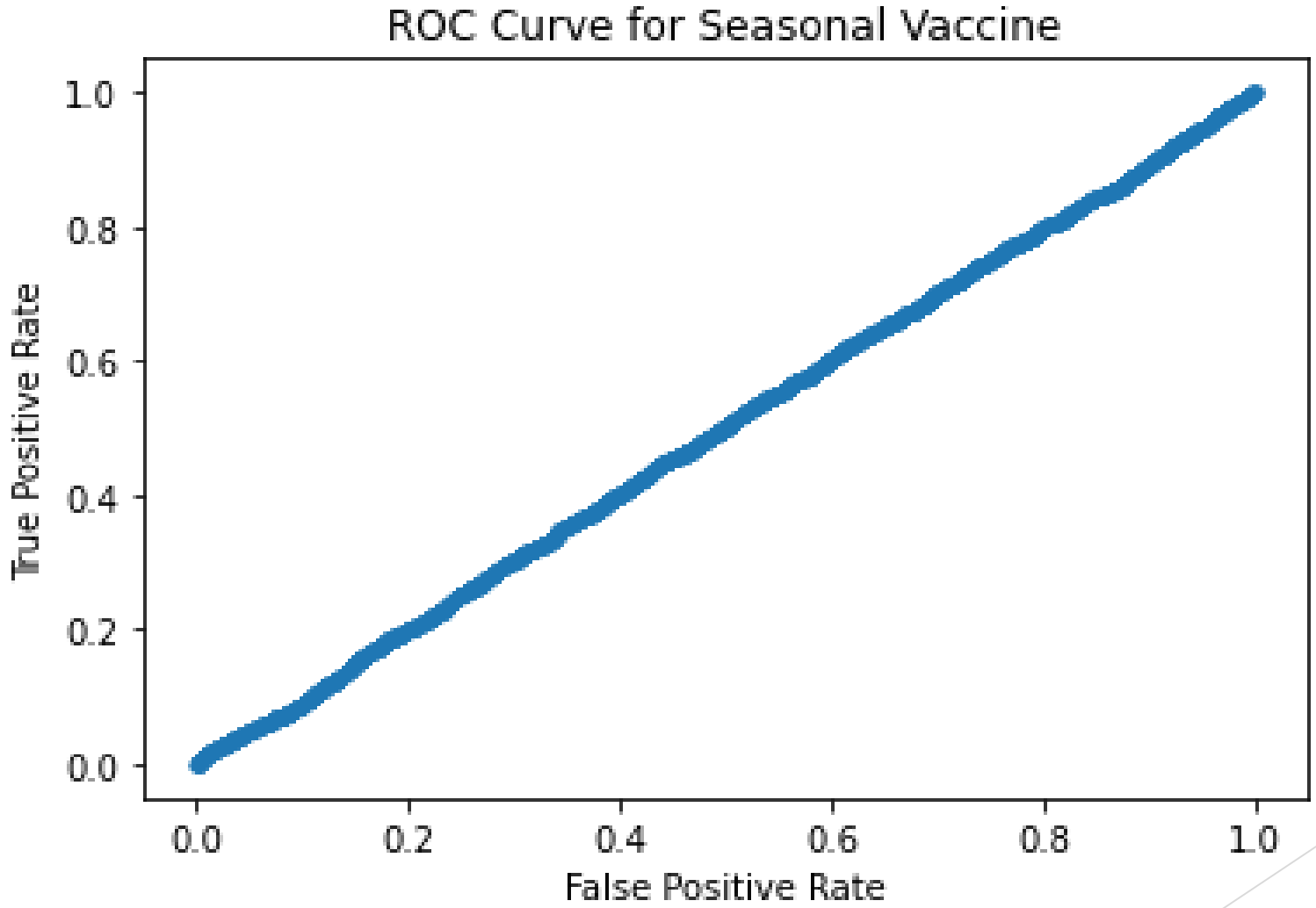
Modeling

- Modeling was done using Logistic regression and Decision tree models. ROC AUC score curve for both vaccines given showed 0.5
- Both vaccines predicted ROC AUC of 0.5 indicating that the models are not effectively distinguishing between classes.
- The features used may not be informative or relevant for predicting vaccine uptake.

Visualisation: ROC Curve for H1N1

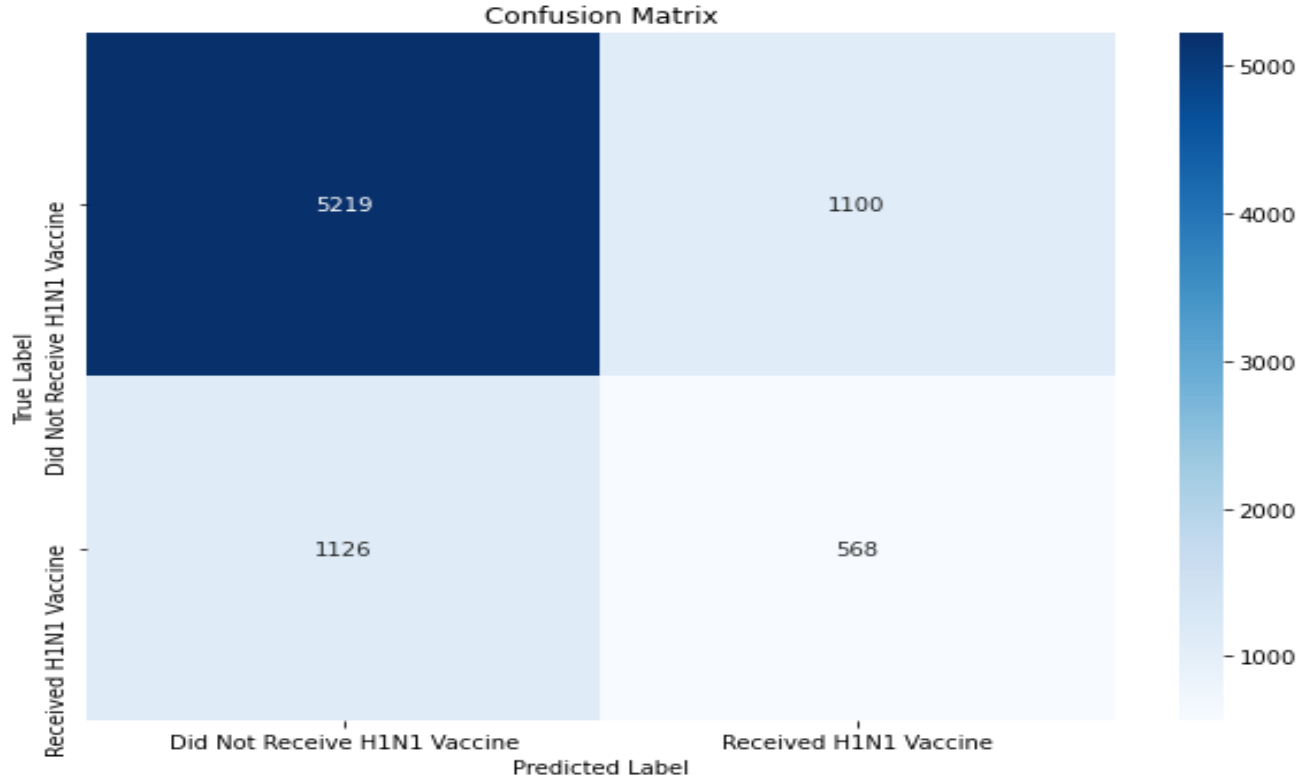


Visualisation: ROC curve for Seasonal vaccine



visualization: Confusion matrix

- Decision tree model was used to evaluate the performance and deciding which models the best



Insights and recommendations

- True Positives (TP): The number of people correctly predicted to have received the H1N1 vaccine. (568)
- True Negatives (TN): The number of people correctly predicted not to have received the H1N1 vaccine. (5219)
- False Positives (FP): The number of people incorrectly predicted to have received the vaccine but did not. (1100)
- False Negatives (FN): The number of people incorrectly predicted not to have received the vaccine but did. (1126)

Cont…..

Precision: 0.34

- Of all the instances predicted as class 1, 34% were actually class 1.
- The model is less accurate at identifying individuals who received the vaccine when it predicts them as such.

Recall: 0.34

- Of all the actual class 1 instances, 34% were correctly predicted as class 1.
- The model misses a significant portion of actual vaccine recipients, indicating low sensitivity for this class.

Cont…

- F1-score: 0.34

- The harmonic mean of precision and recall for class 1.
- The model's performance for class 1 is weak, with a low balance between precision and recall.

Model performance

- The model performs well in identifying class 0(respondents who did not received H1N1 vaccine but poorly in identifying class 1 (respondents who receive vaccine)
- This suggests an imbalance between classes, where the model is biased towards the majority class (class 0).

Recommendation

- Address class imbalance using techniques like resampling (e.g., SMOTE for oversampling the minority class or random underdamping of the majority class), adjusting class weights, or using algorithms designed to handle imbalanced data
- Decision tree classification model performs better than logistic regression

Conclusion:

- Decision tree classification model performs better than logistic regression.