H1N1 & Seasonal Vaccine Adoption- Analysis and Modeling

Data Science approach using Logistic Regression and Decision Tress for Data prediction

Linet Awino Okinyi

2025 Presentation

Introduction

Objective - Predict vaccine Uptake for H1N1 and seasonal vaccines using Machine Learning and make reccomendations

Method used - Logistic regression and dcision Trees

Key Challenges : Class Imbalance , feature selion and model performance

-

About the Data

- Dataset: H1N1 & Seasonal Vaccine Dataset
- Features:
- Behavioral factors.
- Demographics
- Health condition

- Target Variables:
  - h1n1_vaccine: 1 if received, 0 otherwise.
  - seasonal_vaccine: 1 if received, 0 otherwise.

Data Preprocessing

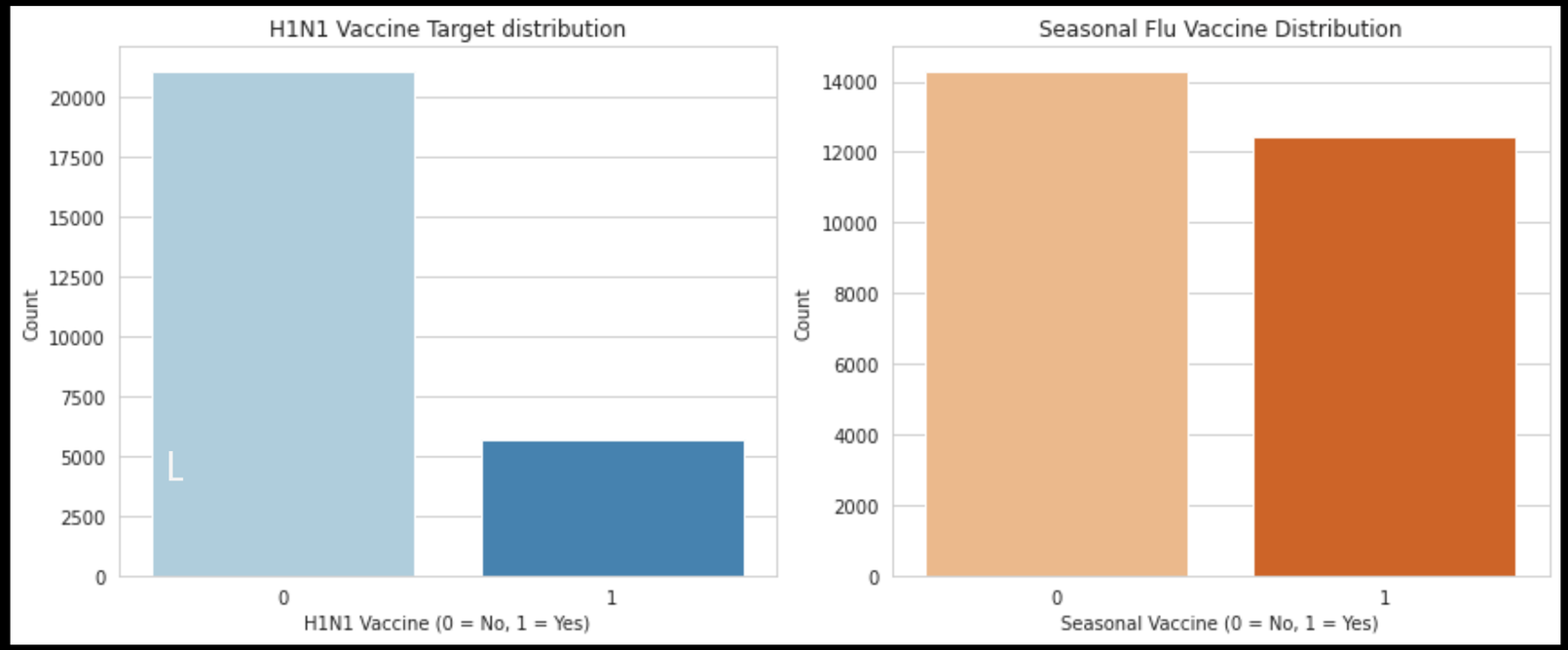
- Dropped highly missing columns

- Imputed missing values with the most frrequent(mode)
- handled data imbalance during modeling class by balancing the class weight.

Feature scaling by standardization

Data Analysis

Proportion of the vaccinated Against Unvaccinated

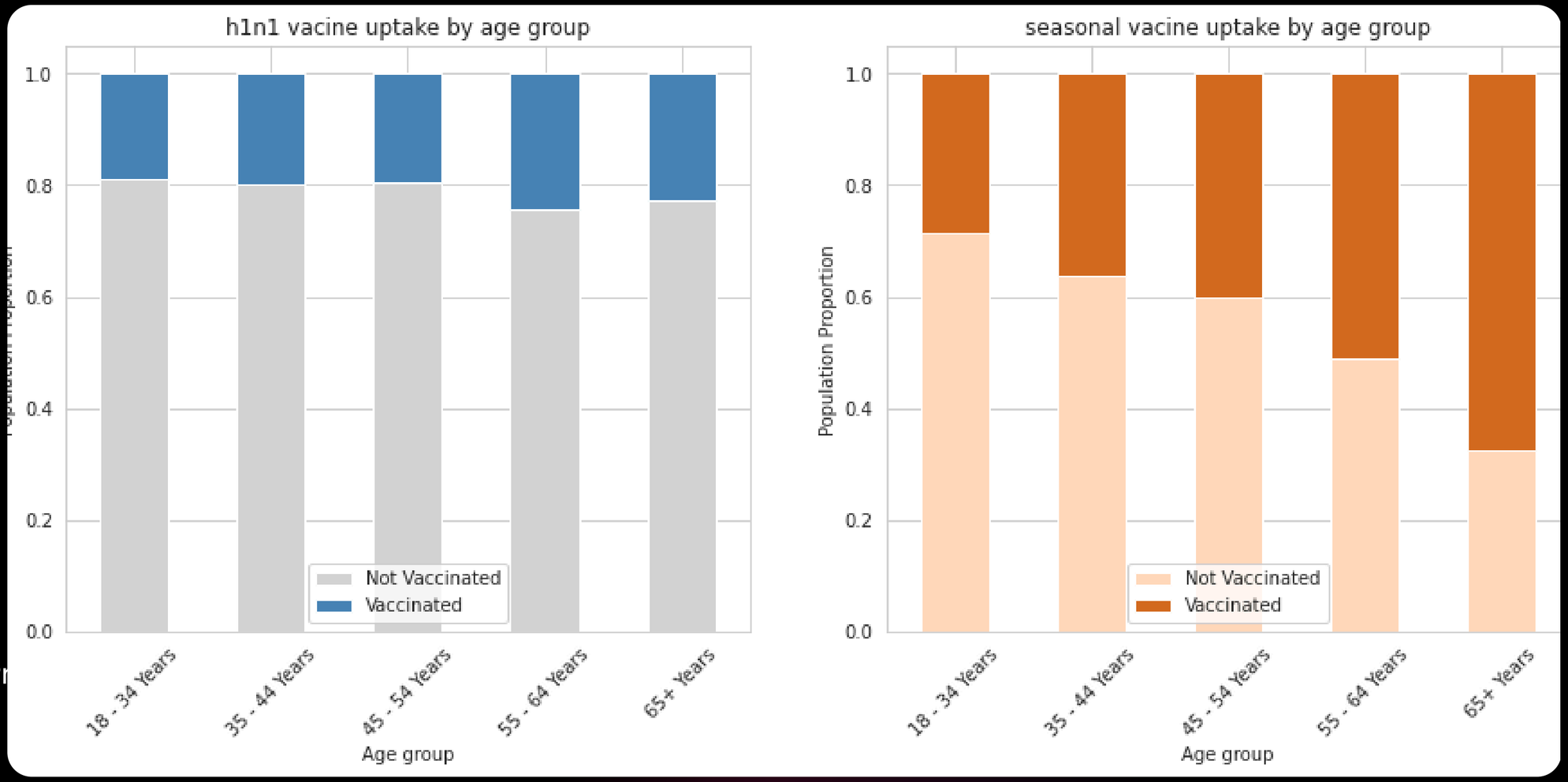


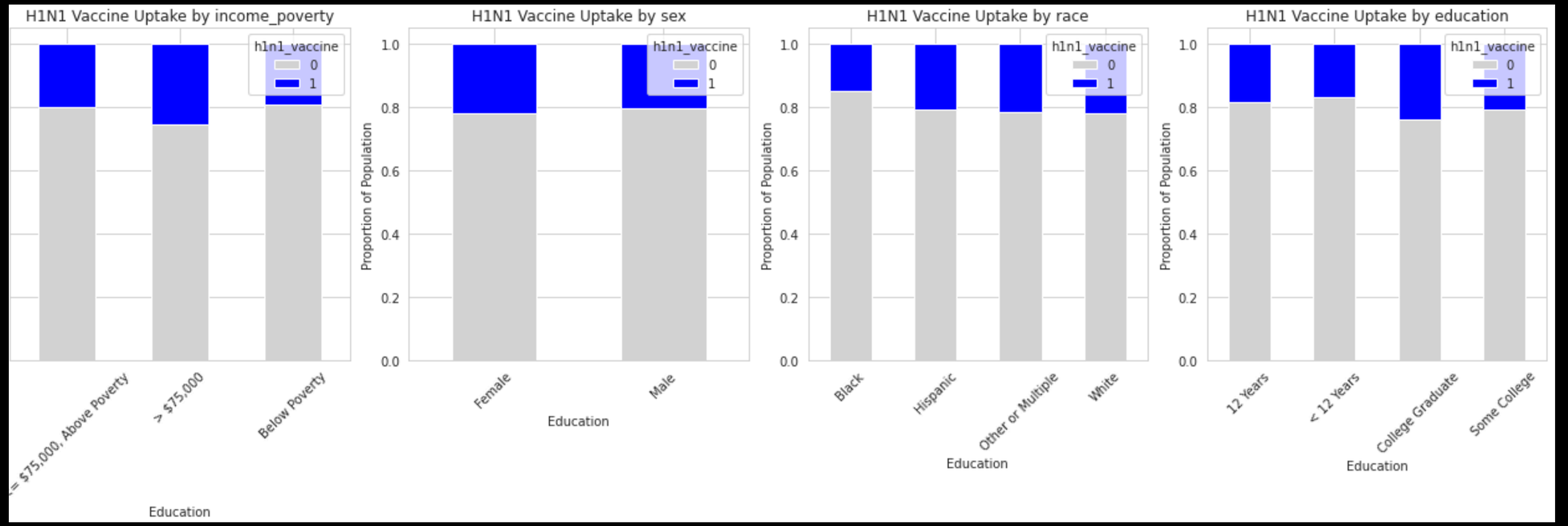
Most people did not take h1n1 vaccine -
A problem of class imbalance

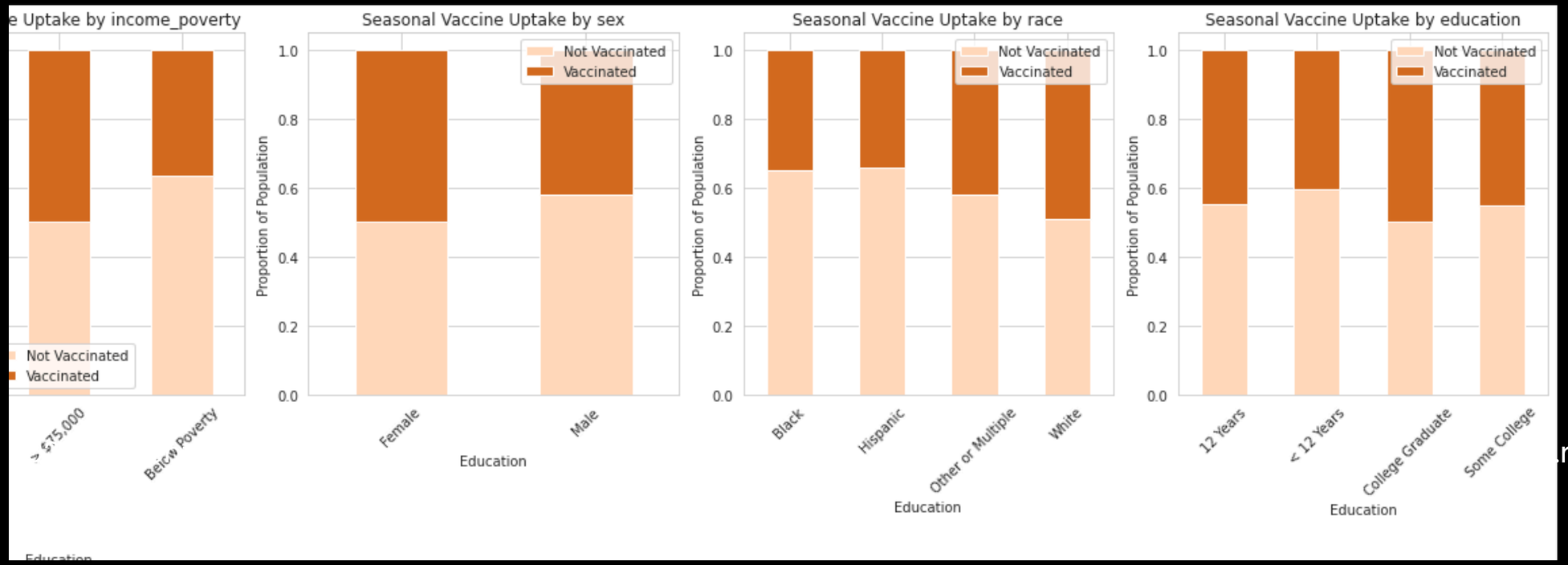
There a good balance between people who took the vaccine and those who did not for seasonal vaccine

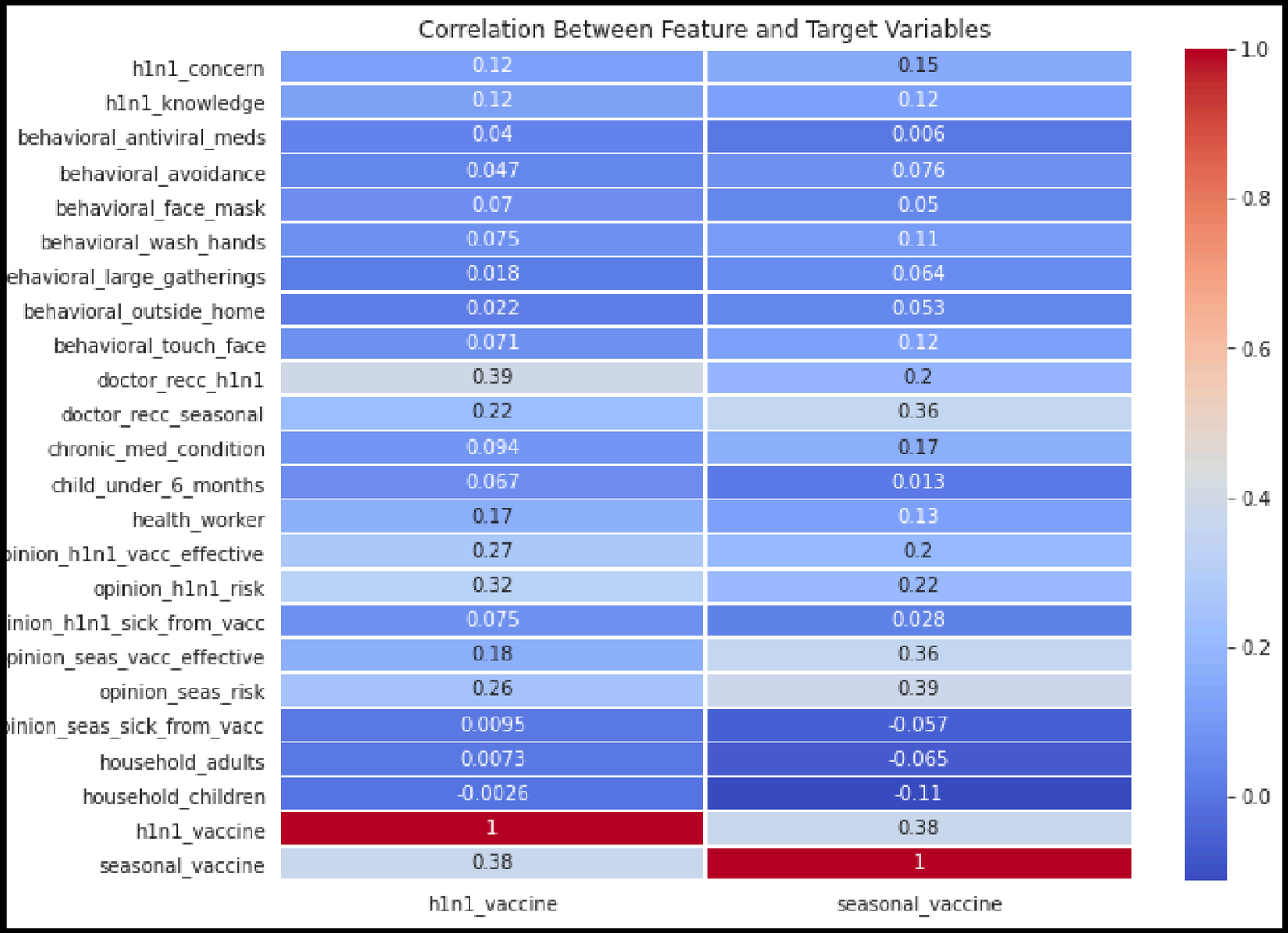
Logistic regression Model

Distribution of ages for vaccinaed and unvaccinated








Correlation Between Feature and Target Variables
h1n1_concern 0.12 0.15
h1n1_knowledge 0.12 0.12
behavioral_antiviral_meds 0.04 0.006
behavioral_avoidance 0.047 0.076
behavioral_face_mask 0.07 0.05
behavioral_wash_hands 0.075 0.11
ehavioral_large_gatherings 0.018 0.064
behavioral_outside_home 0.022 0.053
behavioral_touch_face 0.071 0.12
doctor_recc_h1n1 0.39 0.2
doctor_recc_seasonal 0.22 0.36
chronic_med_condition 0.094 0.17
child_under_6_months 0.067 0.013
health_worker 0.17 0.13
pinion_h1n1_vacc_effective 0.27 0.2
opinion_h1n1_risk 0.32 0.22
inion_h1n1_sick_from_vacc 0.075 0.028
pinion_seas_vacc_effective 0.18 0.36
opinion_seas_risk 0.26 0.39
inion_seas_sick_from_vacc 0.0095 -0.057
household_adults 0.0073 -0.065
household_children -0.0026 -0.11
h1n1_vaccine 1 0.38
seasonal_vaccine 0.38 1
h1n1_vaccine
seasonal_vaccine
1.0
0.8
0.6
0.4
0.2
0.0

H1N1

Train acurracy - 83.13%

Test accuracy - 84.05

AUC scores

Train = 0.8264

test = 0.8295

Poor recall on Positives cases (vaccinated)

Seasonal

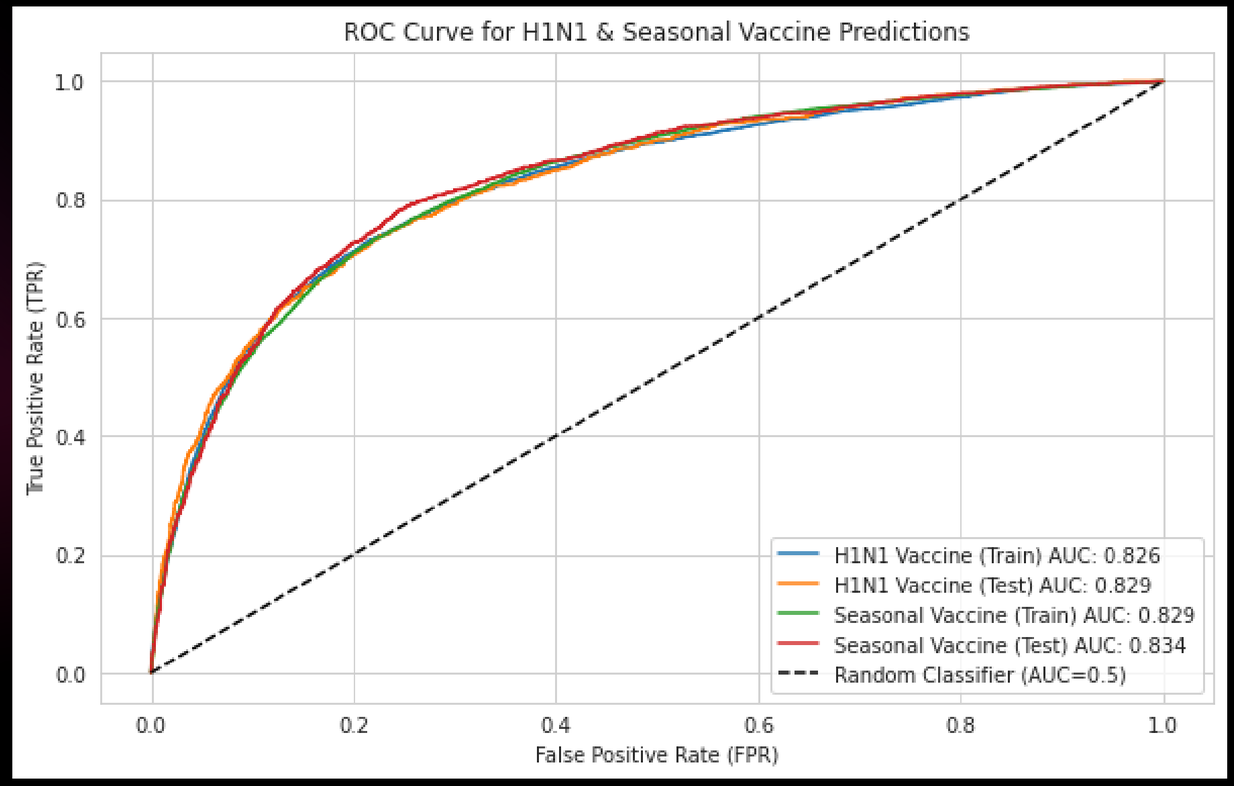
Train Accuracy - 75.79%

test Accuracy 76.68%

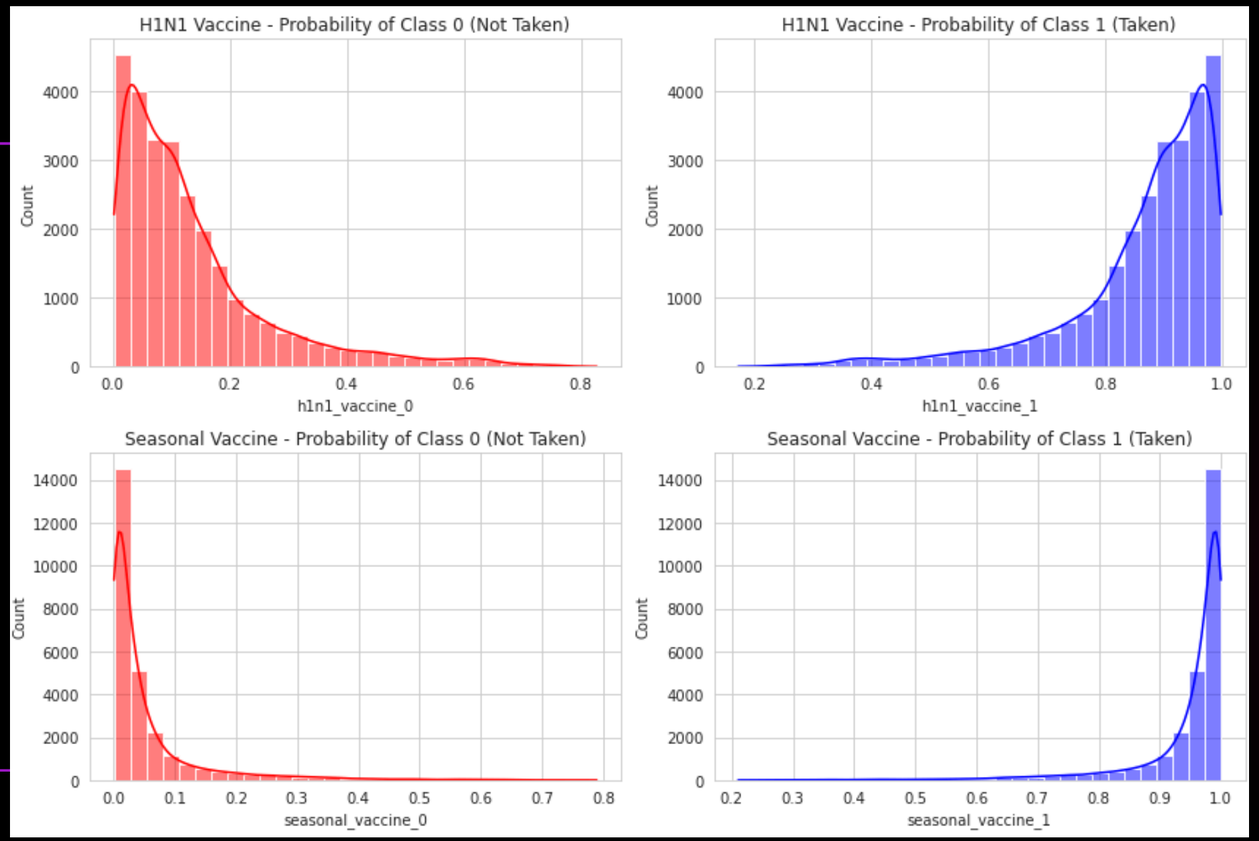
AUC Scores : Train = 0.8289

Test = 0.8340

Logistic Regresson



Vaccine Probability Prediction



The model is making confident predictions

Probabilities are clustering near 0 and 1

Seasonal flu vaccine has more people with high probability predictions that h1n1

Decision Tree

H1N1

accuracy = 75%
Recall:
vaccinated = 0.72
Unvaccinated = 0.76

Seasonal

accuracy = 75%
Recall:
vaccinated = 0.72
Unvaccinated = 0.76

In a Nutshell:

Accuracy dropped from 84% to 75%
The model has a good recall :
The model is making a more balanced prediction

Comparing our 2 models

Key Takeaways:

- Decision Tree provides interpretability but risks overfitting.
- Logistic Regression is stable but struggles with non-linear relationships.

Recommendations

- Healthcare providers (doctors & nurses) should fully paricipate in vaccine sensitization programs

  There is a big gap in public knowlege on vaccine :

  stakeholders should put more emphasis on campaigns educating the public on vaccine safety,efficacy and side effects / risks.

  The younger pupolation should be encouraged to paticipate in vaccination programs

Thank You

FOR YOUR ATTENTION

Questions ???????

2024 Presentation