WanjohiK /
dsc-phase-3-project-v3

`<> Code`   `Issues`   `Pull requests`   `Actions`   `Projects`   `Wiki`   `Security`   `Insights`   `Setting`

Machine Leaning Modeling

☆ **0** stars   ⑂ **0** forks   ⊙ **1** watching   ⑂ Branches   〰 Activity
🏷 Tags

🌐 Public repository

⑂ **1** Branch   🏷 **0** Tags   🔍 Go to file   t   Go to file   +   Add file ▾   Code   ⋯

🟢 **WanjohiK** Update README.md		f93a398 · 1 minute ago 🕘
📄 H1N1 Vaccination Decisions....	Machine Learning	17 minutes ago
📄 Phase 3 Project.ipynb	Machine Learning	17 minutes ago
📄 Phase 3 Project.pdf	Machine Learning	17 minutes ago
📄 README.md	Update README.md	1 minute ago
📄 Tableau Dashboard.pdf	Machine Learning	17 minutes ago

📖 **README**   ✏ ☰

Predicting H1N1 Vaccination Decisions Using Machine Learning: Insights from Demographic, Behavioral, and Attitudinal Factors

Overview

The emergence of infectious diseases has been a significant public health concern worldwide.

Vaccination plays a crucial role in controlling the spread of infectious diseases and mitigating their impact on society

One such outbreak, the H1N1 influenza pandemic of 2009

In response to the 2009 H1N1 pandemic, governments and health organizations launched extensive vaccination campaigns

However, not all individuals chose to receive the vaccine.

Understanding the factors that influence vaccination behavior can help public health authorities design targeted campaigns

Key Business Problem

1. What patterns or trends were observed in the demographic and behavioral characteristics of individuals who received the H1N1 vaccine versus those who did not?(Exploratory Data Analysis)

2. Is there a significant difference in H1N1 vaccine uptake between individuals with different levels of education (e.g., no high school, high school graduate, college degree, etc.)?(Hypothesis testing)

3. Which machine learning models is best for predicting the likelihood of an individual receiving the H1N1 vaccine, incorporating demographic and behavioral data?(Modeling)

4. What were the most important features (demographic, behavioral, and attitudinal) for predicting H1N1 vaccine uptake using machine learning models? (Feature Importance)

Data Understanding

The dataset used in this study is sourced from the National 2009 H1N1 Flu Survey, conducted by the U.S. National Center for Health Statistics.

It consists of responses from individuals across various demographic backgrounds and captures behavioral patterns, health conditions, and personal opinions regarding vaccination.

The dataset contains: 36 feature columns representing different attributes of survey respondents. 1 target variable (h1n1_vaccine) indicating whether an individual received the H1N1 vaccine (binary: 0 = No, 1 = Yes).

Objective 1: Patterns and Trends

The plot clearly shows a significant imbalance in the number of individuals who received the H1N1 vaccine versus those who did not.

The non-vaccinated group (represented by 0) has a much larger count compared to the vaccinated group (represented by 1).

H1N1 Vaccine Uptake
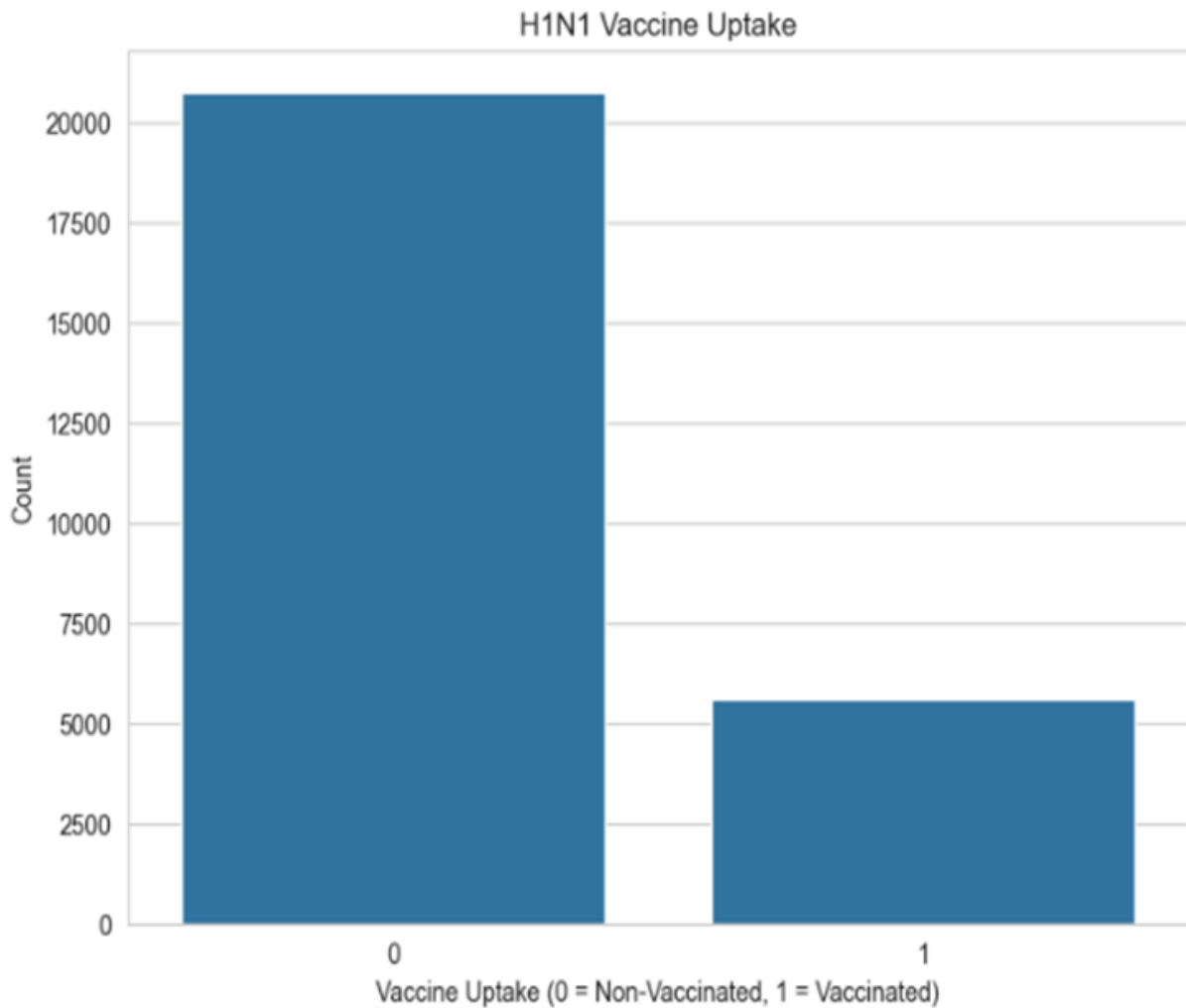
Further Insight

The stacked bar reveals that non-vaccinated individuals are more concentrated in younger age groups, while older age groups have a higher proportion of vaccinated individuals.

This could indicate that age plays a crucial role in vaccine decision-making, with older individuals possibly perceiving a higher risk of H1N1 infection and thus opting for the vaccine.

Age Distribution for Vaccinated vs Non-Vaccinated Individuals
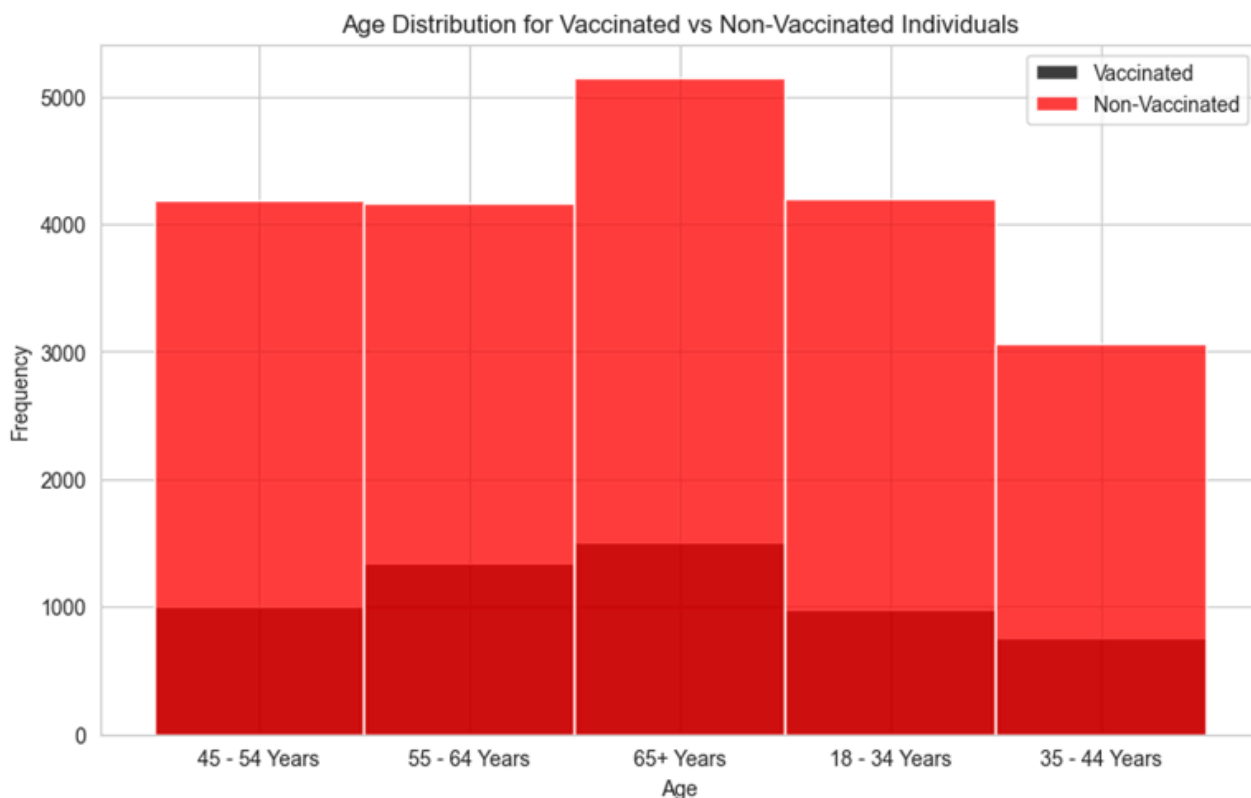
Correlation Heatmap

Behavioral Variables Correlations: Several behavioral variables show moderate correlations with each other, such as:

Mask-wearing, Hand washing, and Avoiding large gatherings have a strong positive correlation with each other (e.g., behavioral face mask with behavioral_wash_hands at 0.34).

This suggests that individuals who engage in one form of preventive behavior are likely to engage in others as well.

Social distancing behaviors (behavioral_large_gatherings and behavioral_outside_home) also show moderate correlations, suggesting that individuals who avoid large gatherings may also reduce contact outside their homes

Objective 2: Hypothesis testing

H_0: There is no significant difference in the uptake of the H1N1 vaccine between individuals with different levels of education.

H_1: There is a significant difference in the uptake of the H1N1 vaccine between individuals with different levels of education

The p-value, which is probability of obtaining the observed results (or more extreme results) under the assumption that the null hypothesis is true, was 5.12e-22 (which is extremely small) is much less than the typical significance level of 0.05 hence rejecting the null hypothesis.

This implies that education level significantly affects vaccine uptake. public health campaigns could benefit from being tailored to different education group

Objective 3: Modeling

This section focused on finding the best model that would predict the likelihood of an individual receiving H1N1 vaccine.

The Metric of focus, according to the competition require was ROC AUC, where a higher value indicates the model was the best.

Several models gave different ROC AUC values

We conclude that, based on the above table, XGboost hyper tuned was the best in predicting whether an individual will take the H1N1 vaccine since its ROC AUC is the highest as shown below

Model	ROC AUC
Base Logistic Regression model	81.8967%
Hyperparameter-Tuned Logistic regression model	81.9024%
Base Decision Tree model	63.7367%
Hyperparameter-tuned Decision Tree model	77.0589%
Base Random Forest model	80.6496%
Hyperparameter-tuned Random-forest model	82.4755
Base XGBoost model	82.7035
Hyperparameter tuned XGBoost model	82.8718

Objective 4: Feature Importance

The feature importance analysis from the XGBoost model highlights that age group (f16), opinion on getting sick from the vaccine, opinion on H1N1 vaccine risk (f14), and H1N1 concern (f0) are the most critical factors.

This suggests that age demographics and risk perception play a significant role in predicting vaccination decisions.

Education (f17), race (f18), and sex (f19) also have substantial importance, indicating that demographic and societal factors strongly influence vaccination behavior.

In contrast, features such as household adults (f23), household children (f24), and behavioral antiviral medication use (f2) have minimal impact.

This suggests that household composition and certain preventive behaviors are not strong predictors in this model.

Feature Importance (XGBoost)

f16 ——————————————————————2398.0
f15 ——————————————————————2328.0
f14 ————————————————— 1975.0
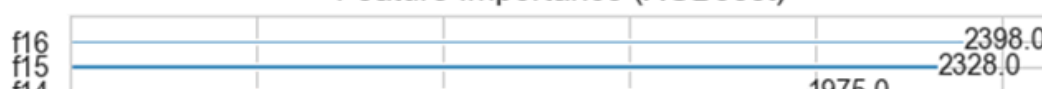
Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

● **Jupyter Notebook** 100.0%